

```

DE HEPARAN SULFATE PROTEOGLYCAN PERLECAN (FRAGMENT) .
GN HSPG2.
OS Homo sapiens (Human)
OC Farkayotis; Metzger; Chondrules; Craniates; Vertebrata; Euteleostomi;
OC Mammalia; Eularchia; Primates; Catartida; Homiidae; Homo.
OX NCBI_TaxID:9606;
RN 11
RA SEQUENCE FROM N.A.
RA Genescope:
RL Submitted (CCF-2000) to the EMBL/Genbank/DBJ databases.
RP 121
RP SEQUENCE FROM N.A.
RP MEDLINE:2055341; PubMed:11101850;
RK Nicolle S., Davoine C.S., Topaloglu H., Gallotio L., Bartal D.,
RK Hechtman P., Ben-Haima C., Hammond B., Girard C., White P.S.,
RK Samson D., Uribeiro J.A., Lehmann-Horn F., Weissbach J.,
RK Hechtman P., Fontaine H.,
RK "Hechtman, the major proteoglycan of basement membranes, is altered in
RK patients with Schwartz-Jampel syndrome (chondrodysplastic myotonia).";
RK Nat. Genet. 26:480-483(2000).
CC -1. SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
CC (LDLR) DOMAIN.
CC EMBL: A144579; CA18543.1;
CC InterPro: IPR000561; EGF_1ike;
CC InterPro: IPR000742; EGF_2;
CC InterPro: IPR003599; Ig;
CC InterPro: IPR003598; Ig_C2;
CC InterPro: IPR003600; Ig_Like;
CC InterPro: IPR003596; Ig_V;
CC InterPro: IPR000034; Laminin_H;
CC InterPro: IPR002049; Laminin_EGF;
CC InterPro: IPR001791; Laminin_G;
CC InterPro: IPR002172; LDL_Receptor_A;
DR EMBL: AF000552; Laminin_H; 3.
DR EMBL: AF000553; Laminin_EGF; 8.
DR EMBL: AF000554; Laminin_G; 3.
DR EMBL: AF000557; LDL_Receptor_A; 4.
DR EMBL: AF013900; SEA; 1.
DR PRINTS: PR00261; LDLRECEPTOR.
DR PRODOM: PD003031; Laminin_H; 3.
DR SMART: SM00181; EGF; 15.
DR SMART: SM00180; EGF_Like; 12.
DR SMART: SM00001; EGF_Like; 7.
DR SMART: SM00409; IG; 22.
DR SMART: SM00408; IG_C2; 22.
DR SMART: SM00406; IG_V; 7.
DR SMART: SM00410; IG_Like; 1.
DR SMART: SM00281; Lamin; 3.
DR SMART: SM00282; Lamin; 3.
DR SMART: SM00192; LDLA; 4.
DR SMART: SM00200; SEA; 1.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS00068; LDLR_2; 4.
KW Glycoprotein.
FT NON_TER
SC SEQUENCE 4370 AA; 466592 MW; F80762C940A657C CRC64;

Query Match
Host Local Similarity 5.8%; Score 102.5; DB 4; Length 4370;
Matches 84; Conservative 39; Mismatches 140; Indels 85; Gaps 14;

```

```

DB 2351 VTIHKKGSLLVRIKHGS-----LLRLYQASPDASGVCVCLASSVPLASVAVTEPA 2406
OY 167 TRP-----PYSKRYVLDHVLKGLSHIRLQNTLDLVNIN-HVIN-----AR 241
DB 2407 GSVVALGVYVVKI-----FSSSSVAEGQILDNLCLVAGQALVAVTIHKKGSLLAR 2450
OY 242 SLKVSPLRI-----YSLIRKLQSMNSQNTMLSLTEHLIGTSYGRGLRVLPES 204
DB 2400 HOVHSIRKLILQVTVADSGVGRVAGSGTDLASVAVTIGQRSGNSGVAVTVHLES 2519
OY 202 NSVDVQLKKLDESAMTANQISVIGOSYRDSFNP-----GVSLRQY 324
DB 2520 SVALANQHTLDNLCLVASQAVITTVYKIKASLPSRHOVWSIRKLP 2567
RESULT 14
OYXMB2 PRELIMINARY HIT 215 AA.
AC OYXMB2;
DT 01-NOV-1999 (TREMBL) 12; Created)
DT 01-MAY-2000 (TREMBL) 14; Last sequence update)
DT 01-OCT-2000 (TREMBL) 19; Last annotation update)
DB 485.1 PROTEIN.
OS Caenorhabditis elegans.
OC Farkayotis; Metzger; Nematoda; Chordata; Rhaditida; Rhaditoida;
OC Rhaditida; Invertebrata; Caenorhabditis.
OX NCBI_TaxID:6249;
RN 111
RA SEQUENCE FROM N.A.
RA Bartlow K.;
RL Submitted (CCF-1998) to the EMBL/Genbank/DBJ databases.
RP 121
RP SEQUENCE FROM N.A.
RP McMurtry A.A.;
RK Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.
RK EMBL: A103267; CA1777.2;
RK EMBL: Z81568; CA1777.2; JOINED.
RK EMBL: Z81568; CA1777.2; JOINED.
RK EMBL: A1032671; CA1777.2; JOINED.
SC SEQUENCE 215 AA; 25209 MW; B6A975C081625390 CRC64;

Query Match
Host Local Similarity 5.7%; Score 100.5; DB 5; Length 215;
Matches 48; Conservative 25; Mismatches 81; Indels 31; Gaps 7;

```


DB	ID	Accession	Name	Length	Date
B01	1211	CACATCCGACGAATAATTGAGCTCATTAATGACCACAGAAGCACTAATTTCAGTCTGTAAACTT		1270	
Gy	1261	cattgcccgaataatgtccatttgcctcgaagaagtlcgaca	1298		
Db	1271	 CATTGCCCGAATAATTCATTTCTCTCCACAAGAACGTCACCA 1308			
RESULT	3				
AK022580					
LOCUS		AK022580	2363 bp mRNA	PRI	24-SEP-2000
DEFINITION		Homo sapiens cDNA FLJ12518 f1s, clone NT2RM2001B05.			
ACCESSION		AK022580			
VERSION		AK022580.1	GI:10434050		
KEYWORDS		oligo coupling; fls (full insert sequence).			
SOURCE		Homo sapiens; fetalocarcinoma cell_line:NT2 cDNA to mRNA. clone_jib:NT2RM2 clone:NT2RM2001B05.			
ORGANISM		Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Hominidae; Homo.			
REFERENCE		Jisoqai,T., Oda,T., Hayashi,K., Sugiama,T., Otsuki,T., Suzuki,Y., Nishikawa,H., Noda,K., Sudano,S., Aotsuka,S., Yoshikawa,Y., Matsumura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Makamatsu,A., Nakamura,Y., Nagahori,I.K., Masuko,Y. and Sakaki,N. NEO human cDNA sequencing project Unpublished (2000)			
TITLE		2 (bases 1 to 2383) Jisoqai,T. and Otsuki,T. Direct Submission Submitted (23-AUG-2000) to the DDBJ/EMBL/Genbank databases. Yako Iseai Helix Research Institute, Genomics Laboratory: 1532-3 Yana, Kisaragi, Osaka 292-0812, Japan (E-mail:j.genom@citru.co.jp, Tel:81-436-3951, Fax:81-438-52-3952) NEO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction: 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.			
FEATURES					
SOURCE		Location/Organisms 1..2383 /organism="Homo sapiens" /db_xref="taxon:9606" /cell_line="NT2" /cell_type="fetalocarcinoma" /clone="NT2RM2001B05" /clone_jib="NT2RM2" /note="cloning vector: pME18SP1.3-mRNA from unduced NT2 neuronal precursor cells." 439..2343 /note="unnamed protein product" /codon_start=1 /db_xref="gi:10434051" /translation="MSLVPTNNYTYTNOLKGGTVAVVGCVKFKRPYSKGNVQGV SVVTVDOTNKLTCLPSNGTEALPIYKKNDIVRHRLKLOYRKKEQGITSSGPV SLTEGLTGAPVIRTSKYTNFTEDIKMYVALKVMASTIMSWTLKLCTVPMK VFIDLITLGRKAAYGASFIARWVGHTRPPSWKVLQD.VLEGDLSIHRLON.T DI.LAYDNHVAWSLKVSFLPILVSLNPKIQSMNSSECPMLSLERHLHGTYSGCI VLPSNSDVDPDKRDLISANTJANDHSIVICOSEHDSPFSQSVALVEVERCOLSI TLLIDHOYLEETPLCALIKOKALOORYIKALKRSYKIDRLPOSYKLIIEPCQLD.OEVI HEGDDIFPKRAKTPDYKADKQNTSKIPKRNKNSVLPYKSGHDELDLDSAPFLIOGTIIHKYGCCSS CLLILEGTLSEICRLSKRNSVLPYKSGHDELDLDSAPFLIOGTIIHKYGCCSS LRKSQNLSNIWKTSMTIPSVAELGVIVLYOVVWTTLIDIGCVLAALMDSPKPKP OI.PASVLEMDIDLOKSDVMDIMDCIPCPIKIDAVFWLCFIRKNVNTGTNPNOLOYOI PDTHVAEVV1"			
BASE COUNT		750 a	456 c	459 g	718 t
ORIGIN					

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

AK001935	2077 bp	mrna	PRI	22-FEB-2000
LOCUS	homo saplens cDNA FLJ11073 fis, clone PLACE1005026.			
DEFINITION	AK001935			
ACCESSION	AK001935			
VERSION	AK001935.1 GI:7023508			
KEYWORDS	oligo capping; fis (full insert sequence).			
SOURCE	homo saplens placenta cDNA to mRNA, clone_lib:PLACE1 clone:PLACE1005026.			
ORGANISM	homo saplens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1 (sites)			
AUTHORS	Isogai,T., Ota,T., Hayashi,K., Shiiyama,T., Otsuki,T., Nishikawa,T., Naqai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y., Makumoto,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahara,K., Masuho,Y. and Sasaki,N.			
TITLE	NEDO human cDNA sequencing project			
REFERENCE	unpublished (2000)			
AUTHORS	2 (bases 1 to 2077)			
TITLE	Isogai,T. and Otsuki,T.			
JOURNAL	Direct Submission			
	Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases, Takao			
	Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yama,			
	Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,			
	Tel:81-438-52-3951, Fax:81-438-52-1952)			
COMMENT	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'- and one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.			

FEATURES	SOURCE	BASE COUNT
Location/Qualifiers		193 4 626 1
1. .2077		
/organism="Homo sapiens"		
/db_xref="taxon:9606"		
/clone="PLAC1005026"		
/clone_lib="PLACE1"		
/library_type="placenta"		
/note="cloning vector: pME18SFLJ"		
131. .2035		
/note="unimad protein product."		
/codon_start=1		
/protein_id="BAA91988.1"		
/db_xref="GI:7023509"		
/translation="MSLVATNYLYTPLNOLKGGTIVNYGVCKVFKPKPPVLSKDTYQSVVTVDQTNVRLTCLLSENGTEALPIYKKNQIVRSHRLQVYKKEQTQIPSSGFA SLTFEGTLAGLIPTRSSYKPFNTEDHKHKEVLRVWASTHMSFTWLKLALADVQMPQYFDITCOLLCKAEVDGASPLIKVWDCTRTTPSWRVLQDVLVFGDLSLTHRLNLTITDLVUNHNVHVARSLKVSFLRILYSLHTPKLOSMSENOTMLSLRPHIIGGTSYGRGIRVLPSNSDNDQLKKDESAITFANGHSIDVLCQSEPDSDSEPSGVSLSLYEVRQQQLSA TLTIDHQLYRTPLPLCA LKOKAPQOYRIRAKRSYKPRFLPSVKHLCKHCLQLQVPHREGDDIIFQDQKATPDVQLQNTSLSLGSLTWTTKNGKGRKVAHVKNNGHLPISNE CLLLKEGCTSLCKLSLNKNSESVIPVRSCHIDLELDLSDPFLQCTLIHHYVCKQKSSFLRSIONLSLVDKTSVTPSSVAEALIVPVLQYVFPVPTTLDGTCGVLKXATLHNSDKPFGTPASVFLMDDDLQKSVDMIDMFPCPGPGLKIDAVYVLECFTRKSYNTNCTGDDNLCGYIIFPTVAEDVI"		657 4 301 5
CDS		

Query Match	80.74%	Score 1048	DB 9	Length 2077
Best Local Similarity	90.84%	Prod. No. 7.5e+268		
Matches 1178	Conservative	0	Mismatches	0
			Indels	120
			Gaps	1
QY	1	atgtcttttgggtccgcagcaatctatataataaccctgaatcaactctaaagatcatt	60	
DB	131	ATGCTCTTTGGTTCAGCAACAATTTATATATATACACCCCTGAATCAACTTAAAGGCTGGT	190	
QY	61	acaatctgcacatgctctctctggtctgaattctcttbaagcccccatctctaaagcaagaa	120	
DB	191	ACAAATGTGCAATGTCTATGCTGTGTTCGAATCTTTTAAGCCCCCATATCTTAAGCAAAAGA	250	

121	QY	121	actgaatatactgctcaatctataactatattatgagccagacaaatctgaaataactaacttgcctcg	180
Db	251	Db		
QY	181	Db	ctctcttaqtgaagaactatgaagccctctccatctaatctatacaaaatctgagatatactctcgc	240
Db	311	Db	ctctctttagtcgaaacattatcgaacgccttcgaataattttatataaaatggagatatgtttcgcg	370
QY	241	QY	ttctccaaqctggaagatctcaatctatacaaaatggagactctcaagatataccaaqctctcgcg	300
Db	371	Db	ttttcacagcgcctgaagatattcaattatataataaaagagagactcagaggtattaccagactctgcgc	430
QY	301	Db	tttgcactctttgacgttttcagacgaacattttgggagagccctctatctatcctcgcgactctcaagc	360
Db	431	Db	aaatattctaacctctcactctctctggaagccctatactcaaacctctcgaactctcaagc	490
QY	361	Db	aaatattctaacctctcactctctctggaagccctatactcaaacctctcgaagcctctcaagcctctcaagc	420
Db	491	Db	aaatatttttaacttttcactatctcagcagaccacaaaatctgtatagaagcctttacgtttgtcgaca	550
QY	421	Db	ctctactcaatactctcaacgctctctggaactctctctcaaaaattctgtctgtctctcaacccaatctcag	480
Db	551	QY	tcctactcaatattctacacgcgtttctggacattttacataaaattctgtgtatcttgcagccaattcgag	610
QY	481	Db	tattcttgcagcctaacctctctcagctctctctctggaagccctaaatctggaagcagcctctctctct	540
Db	611	Db	tatttttcagcctcacttctctcagctctctctctggaagccctaaatctggaagcagcctctctctct	670
QY	541	Db	ctaaaagttatctggaatctgagcccaaggaacccatllctccatctctctggaaggtctctaatcaaacagc	600
Db	671	QY	ctaaagctattggcattggccacagcagcagcagcattttcccatctttggcagagctcttataacaagac	730
QY	601	Db	cttctctcttgcgaaggttaattctaaatctcaatctcagctctcagctctcagctctcagctctcagctct	660
Db	711	QY	ctttctcttctggaaaggttaatttaactctacatctcatagctctatcaaatctggaagcagcatt	790
QY	661	Db	ctatctctcaatctaaacatctctctctctggaaggtctctctctctggaaggtctctctctctctctct	720
Db	791	QY	tttatctctacataaaacataattttatctctctcaacattctctctctctctctctctctctctctct	850
QY	721	Db	atctctctcaagctctcaaaccaatctctcaatctcaatctcaatctcaatctcaatctcaatctcaatct	780
Db	851	QY	atctctatagcctctctataccacaaattctcaatctcaatctcaatctcaatctcaatctcaatctcaat	910
QY	781	Db	ctaaagctctcaatctctcaatctcaatctctcaatctcaatctcaatctcaatctcaatctcaatctca	840
Db	911	QY	tttagactttctatctctcaatctcaatctctcaatctcaatctcaatctcaatctcaatctcaatctca	970
QY	841	Db	atctaaactctctcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatctca	900
Db	971	QY	agtaactctctatctctgagttcaattgaaaaggaatttttagaattctgaaaattttgacagagcatt	1030
QY	901	Db	caatctctcaatctctcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaat	960
Db	1031	QY	cagcattctcagaactttatctctctcaatctcaatctcaatctcaatctcaatctcaatctcaatctca	1079
QY	961	Db	ctctgagctctcaatctctcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatctca	1020
Db	1080	QY	ctctgagctctcaatctctcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatctca	1079
QY	1021	Db	ctctgagctctcaatctctcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatctca	1060
Db	1080	QY	ctctgagctctcaatctctcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatctca	1090
QY	1081	Db	ctctgagctctcaatctctcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatctca	1140
Db	1091	QY	ctctgagctctcaatctctcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatctca	1150
QY	1141	Db	ctctgagctctcaatctctcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatctca	1200
Db	1151	QY	ctctgagctctcaatctctcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatctca	1210
QY	1201	Db	ctctgagctctcaatctctcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatctca	1260

OS Oxytricha trifallax.
 OC Eukaryota; Alveolata; Ciliophora; hypotrichs; Stichotrichida;
 OC Oxytrichidae; Oxytricha.
 OX NCBI_TaxID=5946;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Prescott J.D., Dubois M.L., Prescott D.M.:
 "Oxytricha trifallax micronuclear alpha telomere binding protein
 gene."
 RL Submitted (May-1998) to the EMBL/Genbank/DBJ databases.
 BK EMBL: AF067831; AAC27615.1;
 BK HSSB: P29549; 107C.
 BK InterPro: IPR003415; Telo_bind_alpha.
 BK Pfam: PF02307; Telo_bind_alpha; 1.
 SQ SEQUENCE: 497 AA; 56361 MW; 132C768D6C3A866 CRC64;
 Query Match 8.14; Score 144; DB 5; Length 497;
 Best Local Similarity 21.98; Pred. No. 0.00019;
 Matches 82; Conservative 57; Mismatches 142; Indels 94; Gaps 17;
 QY 7 TNYITPLN-QLKGGTIVNVGVVKKFPPYLSKCTVCGSVTVV-----Q 52
 DB 34 SKVEVVELTKQIVSVQVHVAVVDATPPYKPNQERYTSLKIVDSYIAKKEKCTD 93
 QY 53 TNAVITCLPSGNTVALDITYKNGDIAVFHKLGVYKKEITQG-----TSSGFASLT 106
 DB 94 NSDVATVLYAKRFEDLPDIIHLCDDIHKHATIRLNGQHPNANIPYSSMALPSTDK 153
 QY 107 -----KCTGAPDIPRT-SKYPNFTTDIKIKVEALRVNASPLM-----SLWS 148
 DB 154 KSAIDIEIGGQAT-SDIAPFSHCKNFTPEKKNAGIVQNIKKMAAGVPTGYVASINMF 211
 QY 149 TLKIKGVDPQGYDPLTCLGKAFVGCASPLKVMKCTKTPFSWKLTDHVIKIDLS 208
 DB 212 VALNFAVQKQKIPVAVAKHLDHLEDEYIMKIKDPSAVI YLALFKET 263
 QY 209 HNH-----RLONLTID-----TLVYDNHVAKSLKVGSEFLYSILTKIOSMNSFNO 256
 DB 204 HLKGEVVKIKSATYDDEISTOKKVLISHYNSIVTFTSTSLAR-ELKGNITDRAVER 321
 QY 257 TMLSLFPH-----HGTSYGCGIVLDP-----SNSDIVQLKRLIESANLT 298
 DB 322 AALKQVSLSAVVLTEVDRKHAG-----LVNHSIDLPFHNDSDP--KELSKITF 369
 QY 299 ANQISDVICQSEPPD 313
 DB 370 KQGF--VYTKLEPSD 382
 RESULT 8
 Q45595 PRELIMINARY: PRT; 251 AA.
 AC 045595:
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, last sequence update)
 DT 01-JAN-1999 (TrEMBLrel. 09, last annotation update)
 DE F57C2.3 PROTEIN.
 DE F57C2.3.
 GN Caenorhabditis elegans.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Mammalia; Chordata; Chordata; Chordata;
 OC Rhadaliidae; Peleodermata; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Haynes C.:
 RL Submitted (Nov-1996) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-94150718; PubMed-7906398;
 RA Wilson R., Alnsough K., Anderson K., Haynes C., Herks M.,
 Bonfield J., Hurler J., Connell M., Gopcey T., Cooper J., Toulson A.,
 Craxton M., Dear S., Du G., Durbin K., Fawcett A., Fulton L.,

KA Gardner A., Green P., Hawkins J., Hillier L., Jier M., Johnston L.,
 KA Jones M., Kershaw J., Kirslen J., Laister R., Latreille P.,
 KA Lightning J., Lloyd G., Murray A., Mortimore R., O'Callaghan M.,
 KA Parsons J., Percy C., Ricketts L., Roopar A., Saunders D., Shownkeen R.,
 KA Smardon N., Smith A., Sonhammer E., Staden K., Sultison J.,
 KA Thierrey-Mieg J., Thomas K., Vaudin M., Vaudin K., Waterson K.,
 KA Watson A., Weinstock L., Wilkinson-Sprout J., Woldman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome 111 of C.
 elegans."
 RL Nature 368:32-38(1994).
 BK EMBL: Z83110; CAB0525.1;
 BK SROUNCE 251 AA; 28987 MW; BP2P231001C3B89F CRC64;
 Query Match 7.48; Score 140; DB 5; Length 251;
 Best Local Similarity 25.78; Pred. No. 0.0013;
 Matches 52; Conservative 40; Mismatches 66; Indels 54; Gaps 10;
 QY 160 QYDITCLGKAFVGCASPLKVMK-----TTPPYSKRV----- 196
 DB 25 RYHVALQVSVVETTINGOMLRKVMGRKRGKCAAEERERKLPHTQDSKRYVPPNP 84
 QY 197 LLDIVLEKCLSIDHKLONLTIDILVYDNHVAKSLKVGSEFLYSILTKIOSMNSFNO 256
 DB 85 RICKALFENG-----KLLLEHIVYDHRIGIKNLKNSGDFVALGVNVA-----STRQ 132
 QY 257 TMLSLFPHLPGTSYGGKQVLPDTP--SNSDIVQLKRLIESA-----NLTANQISDV 305
 DB 133 TENGQV--LHGCGASVQKQITTVVDFEHPAPQPKKVEALTEVAVDENPTQGPPE 189
 QY 306 ICOS---EP-DPSFGNSLR 325
 DB 190 VALNVDHEDVQHEALRGVYLK 211
 RESULT 9
 Q9C2K8 PRELIMINARY: PRT; 759 AA.
 AC 09C2K8:
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, last annotation update)
 DE CONSERVED HYPOTHETICAL PROTEIN.
 GN 3H10.40.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Schulte D., Alqn V., Holmset J., Brandt P., Farlman B., Holland K.,
 RA Nyakatura G., Mewes H.W., Mhammedi G.:
 RL Submitted (Feb-2001) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA German Neurospora genome project:
 RL Submitted (Feb-2001) to the EMBL/Genbank/DBJ databases.
 DE EMBL: AL514442; CAC2864.1;
 SQ SEQUENCE: 759 AA; 85218 MW; DC717A5474FF4F73 CRC64;
 Query Match 6.28; Score 111; DB 3; Length 759;
 Best Local Similarity 20.78; Pred. No. 0.29;
 Matches 63; Conservative 50; Mismatches 125; Indels 66; Gaps 12;
 QY 3 LVPATNIVYPPINOL-----KCGTIVNVGVVKKFPPYLSKCTVCGSVTVV-----V 51
 DB 18 LPTST---LTLKALIDDKQAGTSMVNVGVYKADCAVATHTGSDMKCTITSDLSIF 74
 QY 52 QTNVKTCLPSGNTVALDITYKNGDIAVFHKLGVYKKEITQGITSSGFASLTFFCTIG 111
 DB 75 DSSAVTAVAT--RPDABRPVVAQAVIVVLSAKVQKRNKRLITSKITTVGVYVNAAT 132

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection";
 RL Nature 409:685-690(2001).
 DR EMBL: AK013364; BAB28810.1; ..
 DR MGI:1920086; 2810458H16R1k.
 SQ SEQUENCE 113 AA; 12740 MW; 008BB9FC5D0C7088 CRC64;

Query Match 18.2%; Score 324; DB 11; Length 113;
 Best Local Similarity 73.3%; Pred. No. 2.7e-21;
 Matches 63; Conservative 6; Mismatches 17; Indels 0; Gaps 0;
 QY 1 MSIVPATNYIYPLNQLKGGTIVNNGVGVKFKPPYLSKGTGYCSVVTIVDQTNVKTCL 60
 DB 1 MSAPVADSTYPLMLLEGTIVNNGVGVKFKPPYLSKGTGYCSVVTIVDQTNVKTCL 60
 QY 61 LPSGNYEALPIYKNCIDIVRHLKFI 86
 DB 61 LPSRDYDLPVYKVDIVCFQCLAV 86

RESULT 5
 OL3988 PRELIMINARY; PRT: 555 AA.
 AC OL3988;
 DT 01-JAN-1999 (TREMBlrel. 09, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DE IPOTHETICAL 64.1 KDA PROTEIN C26H5.06 IN CHROMOSOME 1.
 GN SPAC26H5.06.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes;
 OX NCBI_TaxID=4896;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Oliver K., Harris D., Bartell B.G., Rajandream M.A., Wood V.;
 RL Submitted (Sep-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z99126; CAB16192.2; ..
 KW Hypothetical protein.
 SQ SEQUENCE 555 AA; 64111 MW; A79DAA95A0C4F803 CRC64;

Query Match 9.4%; Score 167.5; DB 3; Length 555;
 Best Local Similarity 24.4%; Pred. No. 1.8e-06;
 Matches 77; Conservative 54; Mismatches 118; Indels 67; Gaps 15;
 QY 18 KCGTIVNNGVGVKFKPPYLS-KCT-DYCSVVTIVDQT-----NVKLTCLFSGNYEALPI 71
 DB 38 KNTIVNGLGIVLGTIPSSQSLGRTKDWVTIVYLDMDPTDTSISGIGIHLFSGKGNLIPV 97
 QY 72 IYKNGD IVRFHRLKIQVYKKTGTGTSFGFA-----SLTFEGTLCGAPLIPRTS----- 119
 DB 98 IKVCGPILLHUIITLRSYHRTQGLSKDFRYALWPKDFSSNSKDTLCPOPPRLMKTKGDK 157
 QY 120 -----SKYFNFTEDIKHVEALRVWAS-----THMS-PSWTLKLCDDVQPMQYFDLTC 166
 DB 158 EEQFALLLNKINDEQTNKIKNGELLSSARONQTLGSPSVFSLLSQITPHQRCSFYA 217
 QY 167 QLLGKAEVDGASFLKLVNDGTR-----TPPS---WRVLIQDLVGLGDLSHLRLONL 216
 DB 218 QVI-KTWYSDKNFTLYVDYTENELFFPMSPTSSRRW-----GPGF-----RF 261

QY 217 TIDLVDNHHVARS-LKVCSEFLRYSLHTKLOSMHSENQTMLSLFLHIGTGS--YGR 273
 DB 262 SIRCILWDEHDFYCRNTYKGGDYVMKRVTKIDILG-----YLCILHGDSAKRYNM 314
 QY 274 GIRLPESNSDQDLK 289
 DB 315 STEKVDSEPELNEIK 130
 RESULT 6
 Q9FNH7 PRELIMINARY; PRT: 463 AA.
 AC Q9FNH7;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE GBIAAD29059.1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 OX NCBI_TaxID=3702;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE=98069011; PubMed=9405947;
 RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
 RA Tubata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosomes 5. II.
 RT Sequence features of the regions of 1,044,052 bp covered by thirteen
 RT physically assigned pl clones";
 RL DNA Res. 4:291-300(1997).
 DR EMBL: AR006700; BAB08953.1; .. A735DBF109155D7D CRC64;
 SQ SEQUENCE 463 AA; 53630 MW;

Query Match 8.6%; Score 153; DB 10; Length 463;
 Best Local Similarity 23.7%; Pred. No. 2.8e-05;
 Matches 79; Conservative 49; Mismatches 138; Indels 68; Gaps 13;
 QY 23 VNVYGVVKKFPYLSKGTGYCSVVTIVD--QTNVKTCLIFSNYEAALPIYKNGDIVR 80
 DB 25 VNLGIVIVELG---FSNGDCSCETLKIVDPMYSGSLGVKPVKVFART(RDLPKRVESIGDIL 81
 QY 81 FHRUKIOVYKKTGGL---TSSGFASLTGCTIGAPLIPRTSSKYFNFTEDIKHVEAL 116
 DB 82 LSRVKIVLNRKITALCNETTSSFA--LNGKHSVDS(IPYSSPRLFMFODKNLSNL 139
 QY 117 RVWASTIM---SPSWTLKLCDDVQPMQYFDLTCGLGKAEVDSGASFLKLVNDGTRTPP 192
 DB 140 REMMITVYKFDGSGCFTSLK-DLEKGECSNLSCQIVHLSKVYKDRWYLVNDGTE--MP 195
 QY 193 SWRVLIQD---LVLEGLDLSHLRLQNLITDILVYDNIHIVARS-LKVCSEFLRYSLHTKL 248
 DB 196 PCNTLVKSERLPLCVDPEMLPTVNCRKP-----TGSVLR-----I 233
 QY 249 QSMHSENQTMLSLFLHIGTGYCRGIRVLPESNSDQDLKDLIESANLITANQISDVIQ 308
 DB 234 VDRVSEKQA INCLQ-----PCQIVKLL-----NLFFQVNMGLWNATFTTSTKMOYMS 281
 QY 309 SEPDDSPNGVSLRP-----PCWSSVAR 331
 DB 282 REMEAFSPQRFLLTPFLILCMGKPKSPRWNP IAR 315
 RESULT 7
 O76380 PRELIMINARY; PRT: 497 AA.
 AC O76380;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE ALPHA TELOMERE BINDING PROTEIN.

CC NM5 OR NM5.
 CC Parametrium telurectia.
 CC Mitochondrion.
 CC Eukaryota; Alveolata; Ciliophora; Opisthokonta; Pericarya;
 CC Parametrium.
 CC NCBI_TaxID=5888;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN=STOCK 51;
 RC MEDLINE=9017493; PubMed=2308823;
 RA Pritchard A.E., Sellman J.J., Mahalingam K., Sable C.L.,
 RA Venuti S.E., Cummings D.J.,
 RT "Nucleotide sequence of the mitochondrial genome of Parametrium."
 RL Nucleic Acids Res. 18:17-18(1990).
 CC -1- CATALYTIC ACTIVITY: NADH + H⁺ (Oxidation) -> NAD(+) + H₂O.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X15917; CAA34053.1;
 DR PIR: S07744; S07744.
 DR InterPro: IPR001570; Oxidored_q1.
 DR InterPro: IPR001570; Oxidored_q1_N.
 DR Pfam: PF003601; Oxidored_q1.
 DR Pfam: PF00662; Oxidored_q1_N.
 RM Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
 SO SEQUENCE 570 AA; 65196 MW; 2463671710000000 CIRCULAR;
 Query Match 5.0%; Score 89; DH 1; Length 570;
 Best Local Similarity 21.2%; Pred. No. 5.4; Matches 106; Gaps 14;
 Matches 58; Conservative 35; Mismatches 75; Indels 106; Gaps 14;
 QY 9 YTPPLNOLKGTIVNVGVKFFKPYLSKQDYGVVTVVDOTNVLKLLSCNVEA 68
 DB 84 YSPFLLT-LTIGVFNLYTYSYFHEPHISR-----LSLNAPIASMILLVNSGN-- 133
 QY 69 LP11KNDIVYPRKLIKQVKKKEUGTSSGFSNIPFECILGAPIIPRISKYPFNF 126
 DB 134 -----LAVFPGWELIGITS--PELLNMGH-KATTP-KSAFKASFRNKP 174
 QY 127 TEDHKVPEALRVMASTH-----MSP--SWTLKLCDDVPMOYFD 163
 DB 175 SDSAVLIALILYANVIDLNFALINSHLYSEKLGSTPQINSMNLISFC----- 225
 QY 164 LKOLLGKAEVDGASPLKVN--DGTKTPYPSWKVLLDIDLVEGLSHIHKLONLTDL 221
 DB 226 -----LTFAPAFKSAOFCHVWLPDSMEAPVPSAL----- 256
 QY 222 VYDNIIVAKSLKXGSPF--RTY-----SLITKRL 248
 DB 257 -----HSATIVSAGVPLLMKPTPILELSLTPKL 285
 RESULT 15
 PREJ_HUMAN STANDARD; PRT: 654 AA.
 AC Q14829; Q15253; Q9UJH0; Q9NU22;
 DT 20-AUG-2001 (Ref. 40, Created)
 DT 20-AUG-2001 (Ref. 40, Last sequence update)
 DE SERINE/THREONINE PROTEIN PHOSPHATASE WITH EF-HANDS-1 (EC 3.1.3.16)
 DE (PEPF-1) (PROTEIN PHOSPHATASE WITH EF-CALCIUM-BINDING DOMAIN) (PP7C)
 DE (SERINE/THREONINE PROTEIN PHOSPHATASE 7) (PP7C).
 CN PEPF OR PEPF OR PP7C.
 OS Homo sapiens (human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

CC NCBI_TaxID=9606;
 RN 111
 RP SEQUENCE FROM N.A.
 RC TISSUE=Retina;
 RC MEDLINE=97471020; PubMed 9326663;
 RA Sherman P.M., Sun H., Mücke J.P., Williams J., Smalwood P.M.,
 RA Nathans J.,
 RT "Identification and characterization of a conserved family of protein
 RT serine/threonine phosphatases homologous to phospholipid retinal
 RT degeneration C."
 RL Proc. Natl. Acad. Sci. U.S.A. 94:11630-11644(1997).
 RN 121
 RP SEQUENCE FROM N.A.; AND ALTERNATIVE SPLICING.
 RC TISSUE=Retinal brain;
 RC MEDLINE=97348589; PubMed=9415685;
 RA Montell E., Kautli E.J., van der Vosse E., Andolfi G., Mariani M.,
 RA Pace A.A., Consales G.G., den Dunken J.T., Mullaly A., Franco B.,
 RT "A novel human serine/threonine phosphatase related to the phospholipid
 RT retinal degeneration C (rdgC) gene is selectively expressed in sensory
 RT neurons of neural crest origin."
 RL Hum. Mol. Genet. 6:1137-1145(1997).
 RN 131
 RP SEQUENCE FROM N.A.
 RC TISSUE=Retina;
 RC MEDLINE=98104127; PubMed=9440643;
 RA Huang X., Honkaniemi K.E.,
 RT "Molecular cloning, expression, and characterization of a novel human
 RT serine/threonine protein phosphatase, PP7, that is homologous to
 RT phospholipid retinal degeneration C gene product (rdgC)."
 RL J. Biol. Chem. 273:14662-14668(1998).
 RN 141
 RP SEQUENCE OF 1-355 FROM N.A.
 RA Graham D.,
 RT Submitted (Jul-1997) to the EMBL/Genbank/DBJ databases.
 RN 151
 RP SEQUENCE OF 356-653 FROM N.A.
 RA May P.,
 RT Submitted (Mar-2000) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: MAY HAVE A ROLE IN THE RECOVERY OR ADAPTATION RESPONSE
 CC OF PHOTORECEPTORS. MAY HAVE A ROLE IN DEVELOPMENT. MAXIMAL
 CC ACTIVITY IS OBSERVED AT PH 8.0.
 CC CATALYTIC ACTIVITY: A PHOSPHOPROTEIN + H₂O -> A PROTEIN +
 CC ORTHOPHOSPHATE (THIS ENZYME IS SERINE/THREONINE SPECIFIC).
 CC COFACTOR: MAGNESIUM.
 CC -1- ENZYME REGULATION: ACTIVATED BY CALCIUM.
 CC -1- ALTERNATIVE PRODUCTS: 5 ISOFORMS; 1 (SICOM HEK), 1A, 1B, 2 AND 3;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. ISOFORM 3 MAY HAVE NO
 CC FUNCTIONAL SIGNIFICANCE.
 CC -1- TISSUE SPECIFICITY: DETECTED IN RETINA AND RETINAL DERIVED Y-79
 CC RETINOBLASTOMA CELLS. ALSO FOUND IN FETAL BRAIN.
 CC -1- SIMILARITY: BELONGS TO THE PPP FAMILIES OF PHOSPHATASES.
 CC -1- SIMILARITY: CONTAINS 10 DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AF023455; AAH82795.1;
 DR EMBL: X97867; CAA66461.1;
 DR EMBL: AF029777; AAC05025.1;
 DR EMBL: Z94036; CAA40074.1;
 DR EMBL: AL096700; CAB86407.1;
 DR HSSP: P08129; 1F3M.
 DR MIM: 300109;
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR000048; IO.
 DR InterPro: IPR000934; Ser_Lin_phosphatase.
 DR Pfam: PF00016; ehand; 3.

```
*A survey of the Mycoplasma genitalium genome by using random sequencing.*
J. Bacteriol. 175:7918-7930(1993).
-!- FUNCTION: THE ENZYMES WHICH CATALYZE THE REVERSIBLE PHOSPHORYLATIONS OF PYRIMIDINE NUCLEOSIDES ARE INVOLVED IN THE DEGRADATION OF THESE COMPOUNDS AND IN THEIR UTILIZATION AS CARBON AND ENERGY SOURCES, OR IN THE RESCUE OF PYRIMIDINE BASES FOR NUCLEOTIDE SYNTHESIS.
-!- CATALYTIC ACTIVITY: THYMIDINE + PHOSPHATE -> THYMINES + 2-DEOXY-D-RIBOSE 1-PHOSPHATE.
-!- SUBUNIT: HOMODIMER (BY SIMILARITY).
-!- SIMILARITY: BELONGS TO THE THYMIDINE/PYRIMIDINE-NUCLEOSIDE PHOSPHORYLASES FAMILY.
-----
THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce or send an email to license@isb-sib.ch).
-----
EMBL: U39684; AAC1267.1; ...
EMBL: U02191; AAD12476.1; ...
TIGR: MG051;
InterPro: IPRO03262; Anthr_phosphorbsylltransf.
InterPro: IPRO03312; Glycosyltransf_3.
InterPro: IPRO00053; Thymid_phosphatase.
InterPro: IPRO00591; Glycosyltransf_3; 1.
Pfam: PF001864; Glycosyltransf_3; 1.
ProDom: PD001864; Glycosyltransf_3; 1.
ProDom: PD005916; Thymid_phosphatase; 1.
PROSITE: PS00647; THYMID_PHOSPHORYLASE; 1.
Transterase; Glycosyltransferase; Complete proteome.
CONFLICT 392..392 V->I (IN REF. 2)
SEQUENCE 421 AA; 4635 MW; 649CD517CD6F5F62 CRC64;

Query Match          5.0%; Score 49; DB 1; Length 421;
Best Local Similarity 22.1%; Pred. loc. 3.5;
Matches 68; Conservative 48; Mismatches 114; Indels 78; Gaps 15;

QY 14 LNLKGGCT IWNVGVVKPKPPHYLSKGTDYSVVTVVD--GTNKKLTCLLFSG--NYALPL 71
DB   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 7 LNKKQGKALNL-AETNMFVNVLNK-----TLDYD[TAFLMAIFKGMNPNELFL 57
QY 72 EVK----NGD[VRF-HRLKTQVKKTKQTGL-----TSSSGFASLTPECTLGAPI 114
DB   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 58 LTRAMDVTGCHLKFNHHCKISVDKHSTGCGDGKVSLAIHLTLTSLGS----- 105
QY 115 IPTSSSKYEFTEDHKMVAALAVAWASTMSPSWTLKLKDVGPMOYFDLT--COLLGR-- 171
DB   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 106 VARLSRGGLGYTGCTIKDLFAAGV-----KTEITDAQAAACLNKNKCFELQQSK 154
QY 172 --AEVDIASPLAKVMWRTRTPPTSPRAVLIDDGLVLEDLSHIHQNLNTIDLLVDNMIVH 229
DB   : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 155 DIAPDVSVLYGLKDTICTVDNSLI-----CLASSIMSKLIA-----VMNEYIFI 197
QY 230 ARSLKWGSEFIATVSLHTKL-----QSMNSENNMTLSLEPIHGCTSYGRIIVLPENSIV 245
DB   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 198 DLKYGKGAFCCTKKLANELAKLMOSTAKSFRRKSIVKLS--DMNOVIIRKAVGNVEAVEA 256
QY 286 DQLKKDIIE 293
DB   : |||::||:
DB 257 NFLKQDILD 264

RESULT 14
NU5M_PARTE
ID NU5M_PARTE STANDARD: PRT: 570 AA.
AC P15584;
DT 01-APR-1990 (rel. 14, Created)
DT 01-APR-1990 (rel. 14, Last sequence update)
DT 01-FEB-1996 (rel. 33, Last annotation update)
DE NADH+H+OUBIUTONE OXIDOREDUCTASE CHAIN 5 (EC 1.6.5.3).
```

RL Nature 373:752-752(1994).
 RN [2].
 RP CONCEPT.
 RX MEDLINE:92241891; PubMed:1572661;
 NA Cleveland J.-B.:
 KT "Identifying coding exons by similarity search: also-derived and other
 MT potentially misleading protein sequences."?
 RL Genomics 12:838-841(1992).
 RL [3].
 RP ALSO FAMILIES CLASSIFICATION.
 RX MEDLINE:66433009; PubMed:3136422;
 NA Gurevitz Y.:
 KT "The Abu family developed through successive waves of fixation
 MT closely connected with primate lineage history"?.
 RL J. Mol. Evol. 27:194-202(1988).
 RL [4].
 RP ALSO FAMILIES CLASSIFICATION.
 RX MEDLINE:91176815; PubMed:1706701;
 NA Jurka J., Mironov Ievic A.:
 KT "Reconstruct and analysis of human Abu genes"?.
 RL J. Mol. Evol. 32:105-121(1991).
 CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
 CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, A ALU WARNING
 CC CONSENSUS SEQUENCES HAVE BEEN CONSTRUCTED THAT CONTAIN ALL SIX
 CC FRAMES. CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
 CC REPEATS.
 CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
 CC CODON. 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
 CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
 CC POLYPLOIDY OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
 CC ACID SEQUENCES.
 CC -1- CAUTION: ALU REPEITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
 CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
 CC ACTIVE. TRANSCRIPTED BY POL. II. NORMAL TRANSCRIPTS MAY CONTAIN
 CC ALU DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
 CC CLONING LIBRARY. ALSO CONTAIN PARTIAL AND/OR REARRANGED COPIES
 CC LOCATED WITH ALU-DERIVED SEQUENCE IN AN ORIENTATION. ALTHOUGH ALU
 CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
 CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
 CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
 CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
 CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
 CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
 CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
 CC BEING REPORTED.
 CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A POTATIVE PROTEIN SEQUENCE
 CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
 CC PART OF ALU REPEAT MAY HAVE BEEN ACTUALLY INCLUDED IN THE
 CC CODING NUCLEOTIDE SEQUENCE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-stb.ch/announc/cy>
 CC or send an email to license@isb-stb.ch).
 CC -----
 CC EMBL: 014567; ?; NOT_ANNOTATED_CHS.
 CC Hypothetical protein.
 CC DOMAIN 1 96 FRAM: 1.
 CC PT DOMAIN 100 195 FRAM: 2.
 CC PT DOMAIN 199 294 FRAM: 3.
 CC PT DOMAIN 298 393 FRAM: 4.
 CC PT DOMAIN 397 492 FRAM: 5.
 CC PT DOMAIN 496 591 FRAM: 6.
 CC SEQUENCE 591 AA: 63790 MW: 6650395735319095 CHG64;
 CC

DY 253 SINGMPLSEFFHJHCSTSY--GCGCIVLPEPSNDVMDLEKRLDLSAN/TNMOISUICQSE 310
ID 343 ARIIRALITVEFVETIGFHVAOAGLLTCS-----DHVASASQSACITCVSH-----RAR 494

DY 311 PDIPIFGCYSLRHHCKMSVAKSULDAASTS 460
ID 394 XXXFTXIKHSILCHRCWSAVAKSRDLTAASTS 423

RESIDUE_FZ
TCGR2_KATP
ID TCGR2_KATP STANDARD: PR1 567 AA.
AC FARAAR;
DE 01-08-1994 (Re-l. 40; Created)
DE 01-08-1994 (Re-l. 40; Last sequence update)
DE 20-AUG-2001 (Re-l. 40; Last annotation update)
DE TGF-BETA RECEPTOR TYPE II PRECURSOR (Fig 2.7.1.47) (TYPE_2) (TGF BETA
ID TYPE II RECEPTOR).
CD TRPBK2.
DS Kallikrein-associated (Krat).
OS Eukaryotic Metazoa; Chordata; Vertebrate; Mollusca; Mollusciformia;
OC Mammalia; Eutheria; Rodentia; Sciurodonta; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
II
BP SEQUENCE FROM N.A.
NC STRAIN-SPECIFIC-DIMINISHES WITH SUBCELLULAR;
KA MEDLINE:9432151B; PubMed:8485458;
EA Struchiner K., Lewis R.A., Matthews L.S., Vale W.W.;
RT "Molecular characterization of rat transforming growth factor-beta
BT type II receptor";
RL Biochem. Biophys. Res. Commun. 191:790-795(1993).
I21
NP SEQUENCE FROM N.A.
RX MEDLINE:9408794B; PubMed:8467545;
EA Choi R.E., Kim I.H., Huang Y.C., Lee J.-young H.J.;
RA "Rat mesangial cell hypertrophy as response to transforming growth
FI factor-beta 1.";
RL Kidney Int. 44:648-658(1993).
CC -1- FUNCTION: TYPE I/TYPE II TGF-BETA RECEPTORS PROM AN HETEROMERIC
COMPLEX AFTER BINDING TGF-BETA AT THE CELL SURFACE AND ACT AS
CC SIGNAL TRANSDUCERS.
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN + ADP + A PHOSPHORYLATED
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- PMR: PHOSPHORYLATED ON A SER/THR RESIDUE IN THE CYTOPLASMIC
CC DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC TGB RECEPTOR SUPERFAMILY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/announcement/
or_send_an_email_to_license@isb-sib.ch](http://www.isb-sib.ch/announcement/or_send_an_email_to_license@isb-sib.ch)).
EMBL: L09653; AAAA2337..1..?
DB EMBL: S67770; AAA2352..??..
DR PIKE: J06059; J06059.
DR HSHP: P00518; JP0R.
DR InterPro: IPRO00472; Activin-like.
DR InterPro: IPRO00719; Biol_kinase.
DR InterPro: IPRO0290; Ser/thr_kin_acctc.
DR Pfam: PF00069; Kinase_1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP_1.
DR PROSITE: PS00108; PROTEIN_KINASE_SF_1.
DM Receptor; Transferrin; Soluble/homocysteine-protein kinase; ATP-binding;
KW Transmembrane; Phosphorylation; Glycoprotein; Signal.
FM SIGNAL. POTENTIAL.
FM CHAIN 24 567 TGF-BETA RECEPTOR TYPE II
FM DOMAIN 24 166 EXTRACELLULAR (POTENTIAL);
FM TRANSMEM 167 187 POTENTIAL.

FT	DOMAIN	360	403	LDL-RECEPTOR CLASS A 4.	FT	DISULFID	1295	1304	BY SIMILARITY.
FT	DOMAIN	404	504	IG-LIKE C2-TYPE DOMAIN 1.	FT	DISULFID	1307	1322	BY SIMILARITY.
FT	DOMAIN	521	530	LAMININ EGF-LIKE 1 (N-TERMINAL).	FT	DISULFID	1563	1572	BY SIMILARITY.
FT	DOMAIN	531	730	LAMININ DOMAIN IV 1 (DOMAIN III A).	FT	DISULFID	1565	1579	BY SIMILARITY.
FT	DOMAIN	731	763	LAMININ EGF-LIKE 1 (C-TERMINAL).	FT	DISULFID	1582	1591	BY SIMILARITY.
FT	DOMAIN	764	813	LAMININ EGF-LIKE 2.	FT	DISULFID	1594	1610	BY SIMILARITY.
FT	DOMAIN	814	871	LAMININ EGF-LIKE 3.	FT	DISULFID	1613	1628	BY SIMILARITY.
FT	DOMAIN	879	923	LAMININ EGF-LIKE 4 (INCOMPLETE).	FT	DISULFID	1615	1638	BY SIMILARITY.
FT	DOMAIN	924	933	LAMININ EGF-LIKE 5 (N-TERMINAL).	FT	DISULFID	1641	1650	BY SIMILARITY.
FT	DOMAIN	934	1125	LAMININ DOMAIN IV 2 (DOMAIN III B).	FT	DISULFID	1653	1668	BY SIMILARITY.
FT	DOMAIN	1126	1158	LAMININ EGF-LIKE 5 (C-TERMINAL).	FT	DISULFID	1792	1839	BY SIMILARITY.
FT	DOMAIN	1159	1208	LAMININ EGF-LIKE 6.	FT	DISULFID	1886	1932	BY SIMILARITY.
FT	DOMAIN	1209	1265	LAMININ EGF-LIKE 7.	FT	DISULFID	1976	2021	BY SIMILARITY.
FT	DOMAIN	1275	1324	LAMININ EGF-LIKE 8.	FT	DISULFID	2073	2118	BY SIMILARITY.
FT	DOMAIN	1325	1334	LAMININ EGF-LIKE 9 (N-TERMINAL).	FT	DISULFID	2170	2215	BY SIMILARITY.
FT	DOMAIN	1335	1529	LAMININ DOMAIN IV 3 (DOMAIN III C).	FT	DISULFID	2268	2313	BY SIMILARITY.
FT	DOMAIN	1530	1562	LAMININ EGF-LIKE 9 (C-TERMINAL).	FT	DISULFID	2365	2413	BY SIMILARITY.
FT	DOMAIN	1563	1612	LAMININ EGF-LIKE 10.	FT	DISULFID	2456	2506	BY SIMILARITY.
FT	DOMAIN	1613	1670	LAMININ EGF-LIKE 11.	FT	DISULFID	2554	2599	BY SIMILARITY.
FT	DOMAIN	1677	1771	IG-LIKE C2-TYPE DOMAIN 2.	FT	DISULFID	2641	2686	BY SIMILARITY.
FT	DOMAIN	1772	1865	IG-LIKE C2-TYPE DOMAIN 3.	FT	DISULFID	2831	2876	BY SIMILARITY.
FT	DOMAIN	1866	1954	IG-LIKE C2-TYPE DOMAIN 4.	FT	DISULFID	2917	2962	BY SIMILARITY.
FT	DOMAIN	1955	2049	IG-LIKE C2-TYPE DOMAIN 5.					
FT	DOMAIN	2050	2148	IG-LIKE C2-TYPE DOMAIN 6.					
FT	DOMAIN	2149	2244	IG-LIKE C2-TYPE DOMAIN 7.					
FT	DOMAIN	2245	2343	IG-LIKE C2-TYPE DOMAIN 8.					
FT	DOMAIN	2344	2436	IG-LIKE C2-TYPE DOMAIN 9.					
FT	DOMAIN	2437	2532	IG-LIKE C2-TYPE DOMAIN 10.					
FT	DOMAIN	2533	2619	IG-LIKE C2-TYPE DOMAIN 11.					
FT	DOMAIN	2620	2720	IG-LIKE C2-TYPE DOMAIN 12.					
FT	DOMAIN	2721	2809	IG-LIKE C2-TYPE DOMAIN 13.					
FT	DOMAIN	2810	2895	IG-LIKE C2-TYPE DOMAIN 14.					
FT	DOMAIN	2896	2980	IG-LIKE C2-TYPE DOMAIN 15.					
FT	DOMAIN	2981	3130	LAMININ G-LIKE 1 (GLOBULAR DOMAIN V A).					
FT	DOMAIN	3049	3241	EGF-LIKE 1.					
FT	DOMAIN	3304	3495	EGF-LIKE 2.					
FT	DOMAIN	3558	3705	LAMININ G-LIKE 2 (GLOBULAR DOMAIN V B).					
FT	SITE	65	67	HEPARAN SULFATE (POTENTIAL).					
FT	SITE	71	73	HEPARAN SULFATE (POTENTIAL).					
FT	SITE	76	78	HEPARAN SULFATE (POTENTIAL).					
FT	SITE	3613	3617	MEDIATES MOTOR NEURON ATTACHMENT (POTENTIAL).					
FT	DISULFID	199	212	BY SIMILARITY.					
FT	DISULFID	206	225	BY SIMILARITY.					
FT	DISULFID	219	234	BY SIMILARITY.					
FT	DISULFID	285	297	BY SIMILARITY.					
FT	DISULFID	292	310	BY SIMILARITY.					
FT	DISULFID	304	319	BY SIMILARITY.					
FT	DISULFID	325	337	BY SIMILARITY.					
FT	DISULFID	332	350	BY SIMILARITY.					
FT	DISULFID	344	359	BY SIMILARITY.					
FT	DISULFID	368	381	BY SIMILARITY.					
FT	DISULFID	375	394	BY SIMILARITY.					
FT	DISULFID	388	403	BY SIMILARITY.					
FT	DISULFID	428	470	BY SIMILARITY.					
FT	DISULFID	764	773	BY SIMILARITY.					
FT	DISULFID	766	780	BY SIMILARITY.					
FT	DISULFID	783	792	BY SIMILARITY.					
FT	DISULFID	795	811	BY SIMILARITY.					
FT	DISULFID	814	829	BY SIMILARITY.					
FT	DISULFID	816	839	BY SIMILARITY.					
FT	DISULFID	842	851	BY SIMILARITY.					
FT	DISULFID	854	869	BY SIMILARITY.					
FT	DISULFID	1159	1168	BY SIMILARITY.					
FT	DISULFID	1161	1175	BY SIMILARITY.					
FT	DISULFID	1178	1187	BY SIMILARITY.					
FT	DISULFID	1190	1206	BY SIMILARITY.					
FT	DISULFID	1209	1224	BY SIMILARITY.					
FT	DISULFID	1211	1234	BY SIMILARITY.					
FT	DISULFID	1237	1246	BY SIMILARITY.					
FT	DISULFID	1249	1263	BY SIMILARITY.					
FT	DISULFID	1275	1287	BY SIMILARITY.					
FT	DISULFID	1277	1293	BY SIMILARITY.					

Query Match

5.2%; Score 92.5; DB 1; Length 3707;

Best Local Similarity 19.1%; Pred. No. 38;

Matches 72; Conservative 56; Mismatches 141; Indels 107; Gaps 15;

QY	49	IVDOTNVKLTCLFSGNYEALPIIYKNGDIVRFH-----RLKIQVYKKETOG-----I	96
DB	2063	VAEQOTLDLNCVV-PGHAHAQVTHMKRGGSGLPTHGTHGSLRLYQVSSADSGEYVCSVL	2121
QY	97	TSSG-----FASLTPECTLGAPITPTSSKYENFTTDEHKMVEALRVMASTH	143
DB	2122	SSSGPLASVLVSITPAANVHIPTGVVPIRIETSSSRVAEGOTLDLSCV----VPCQAH	2177
QY	144	MSPSW-----TLKLKCDVQPMQYEDLTCLLCKAEVDCASFLAAVWATR	198
DB	2178	AQVTHMKRGGSGLPACHOVHGHMLRLNRVSPADSGEYVCSVLTLEASGLT	2437
QY	189	TPPFSWRVLTDLVLEGLDSIIHHLNLTIDLVYDN-HVIV-----ARSLKVG	237
DB	2238	SP1PA-PGLAQPVYTESSSSHLTGQTVDLKCVVPQAHQVTHMKRGSSLPARIHQTHGS	2296
QY	238	FLRLYSLHTK-----LOSMMSENQTMLSLEFHLGGTSYGRGIRVLPSNSIUDQL	288
DB	2297	LRLYQLSPADSGEYVGVAGSSHPHEASPKL-----TVPSSONSSFR	2341
QY	289	KK----DLESAMITANQHSV-----ICSEPPDDSP---PNC--VSLR	322
DB	2342	RSPVISIEPPSSSTVOOCODASPKCLIHFGAMPKVEWKIRDOULEDRVHISPRGSIITIV	2401
QY	323	PGCWSSVARSRLAAS	338
DB	2402	APGPATMEPTACVASN	2417

RESULT 11

ALU1_HUMAN

ID ALU1_HUMAN STANDARD; PRT; 591 AA.

AC P39188;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DE 20-AUG-2001 (Rel. 40, Last annotation update)

DE ALU SUBFAMILY J SEQUENCE CONTAMINATION WARNING ENTRY.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN 11

RP SEQUENCE FROM N.A.

RX MEDLINE=95021758; PubMed=7935834;

RA Claverie J.-M., Mukalowski W.;

RT "Alu alert.";

EMBL: AP002551: BAB3871.1: -
 DK PIR: J50349: J50349:
 DK EcoGene: EC11094: sbcd.
 DK InterPro: IPR003701: DNA_repair.
 DK InterPro: IPR000934: Ser_Like_phosphatase.
 DK Pfam: PF0579: 287: 1094: 1.
 DK Hydrus: Nucleotide: Exonucleases: Complete: Proteome.
 DK SEQUENCE: 400 AA: 4473 MW: 488940891986126 CAC64:

Query Match 5.3% Score 93.5; DR 1; Length 400;
 Best Local Similarity 21.2%; Pred. No. 1.3;
 Matches 69; Conserved 38; Mismatches 119; Indels 99; Gaps 14;

QY 34 PPVLSKRDYGVVTVVQTVWVKTCLPSCNVEALPIYKNDLVPHIKLGVYKKEP 93
 11
 146 PPSVAF-PLVNMVAVNILOOTGCHLVVL-ACNIDSVATLANSNDIMAF-----LNT 102
 QY 94 QGTSNGFASLTF--ECPICADLP-----RTSKRTFTTEHKKVVALRWAS 141
 11
 146 TVVASAGHAWOILDRKCTPKCAVADLPPLADKDIILISQACINCEKQCHLAAITVYQ 162
 QY 142 THMSPTWTLKELGVDPQVYELTQGLGKAVIKAS-ETLKVMEKPTPTSMVEL 198
 11
 146 QHYADA--CKLRGIDPLPI-----ATGCHITVYGASNSDAVDIYICGLADPTAQNTP 214
 QY 199 QDLVLECHLSHIDKQNTETDILVYDNIHVAKSLKVGSLRLYSILTRLOSMSNENGT 258
 11
 146 ALVLA--LGHIDHQAQ-----JICMEHVRVCGS-----P 241
 QY 259 LSLFTHLIGTSYGRGIRVLPESNSDYDQ-----LKKDLSANLTANQSD 304
 11
 146 LPLSLDECGKSKRY---VILVTFSSCKLSESVNINLVITGVMAVLKGLASITPAQLQWRD 298
 QY 305 V1COSEPDSPNGVSLAPPQMSV 329
 11
 146 299 -----VSOEPPVWLI 309

RESULT 10
 PGM_MOUSE STANDARD; PRT: 3707 AA.
 AC Q05793;
 DT 01-NOV-1995 (Ref. 32, Created)
 DT 01-NOV-1995 (Ref. 32, Last sequence update)
 DT 20-AUG-2001 (Ref. 40, Last annotation update)
 DT BASEMENT MEMBRANE-SPECIFIC HEPARAN SULFATE PROTEOGLYCAN CORE
 DT PROTEIN PRECURSOR (HSPG) (PERLECAN) (PLC).
 DT HSPG2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 ON NCBI_TaxID:10090;
 RX MEDLINE:92078153; PubMed:1744087;
 RA Yamada Y., Hassell J.R.;
 RA Nounan D.M., Poller A., Valente P., Cai S., Horian E., Sasaki M.,
 RA "The complete sequence of perlecan, a basement membrane heparan
 RA sulfate proteoglycan, reveals extensive similarity with laminin A
 RA chain, low density lipoprotein receptor, and the neural cell adhesion
 RA molecule."
 RT J. Biol. Chem. 266:22939-22947(1991).
 RN [2]
 RS SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A. AND PARTIAL SEQUENCE.
 RX MEDLINE:89034110; PubMed:2972708;
 RA Nounan D.M., Horian E.A., LeDuc S.R., Wong J.G., Sasaki M.,
 RA Yamada Y., Hassell J.R.;
 RT "Identification of cDNA clones encoding different domains of the
 RT basement membrane heparan sulfate proteoglycan."
 RL J. Biol. Chem. 263:16379-16387(1988).
 CC -!- FUNCTION: THIS PROTEIN IS AN INTEGRAL COMPONENT OF BASEMENT

CC MEMBRANE. IT IS RESPONSIBLE FOR THE FIXED NEGATIVE ELECTROSTATIC
 CC CHARGE AND IS INVOLVED IN THE CHARGE-SELECTIVE ULTRAFILTRATION
 CC PROPERTIES. IT INTERACTS WITH OTHER BASEMENT MEMBRANE COMPONENTS
 CC SUCH AS LAMININ AND COLLAGEN TYPE IV AND SERVES AS AN ATTACHMENT
 CC SUBSTRATE FOR CELLS.
 CC -!- SHOWN: PERLECAN HAS A STRONG TENDENCY TO AGGREGATE IN
 CC DIMERS OR SPICULATE STRUCTURES.
 CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANE.
 CC -!- PMS: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
 CC AND O-LINKED GLYCOSACCHARIDES.
 CC -!- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
 CC -!- SIMILARITY: CONTAINS 10.5 LAMININ EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 3 LAMININ DOMAINS IV.
 CC -!- SIMILARITY: CONTAINS 15 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.
 CC
 CC THIS SWISS PROTEIN IS COPYRIGHTED. IT IS PRODUCED THROUGH A COLLABORATION
 CC BETWEEN THE SWISS INSTITUTE OF BIOFORMATICS AND THE EMBL OUTSTATION
 CC AT THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
 CC USE BY NON-PATENT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
 CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
 CC ENTERPRISES REQUIRES A LICENSE AGREEMENT (See <http://www.ebi.ac.uk/amboss/>
 CC or send an email to license@ebi.ac.uk).
 CC
 DR EMBL: M77174: AAA49911.1: -
 DR EMBL: J04054: AAA30899.1: -
 DR EMBL: J04055: AAA49921.1: -
 DR HSSP: P01301: IAA1.
 DR MGI: MGI:96257: Hsp92.
 DR InterPro: IPR000561: EGF-1-like.
 DR InterPro: IPR000742: EGF_2.
 DR InterPro: IPR001438: EGF_11.
 DR InterPro: IPR003006: IG_MHC.
 DR InterPro: IPR003598: IG_C2.
 DR InterPro: IPR002172: LDL_receptor_A.
 DR InterPro: IPR000034: Laminin_E.
 DR InterPro: IPR002049: Laminin_EGF.
 DR InterPro: IPR001791: Laminin_G.
 DR InterPro: IPR000082: SEA.
 DR Pfam: PF00047: Ig_15.
 DR Pfam: PF00052: Laminin_E_3.
 DR Pfam: PF00053: Laminin_EGF_8.
 DR Pfam: PF00054: Laminin_G_3.
 DR Pfam: PF00057: Ldl_receptor_a_4.
 DR Pfam: PF01390: SEA_1.
 DR PRINTS: PR00010: EGFRAQD.
 DR PRODOM: PD003031: Laminin_E_3.
 DR SMART: SM00180: EGF_Lam_7.
 DR SMART: SM00001: EGF_11ke_6.
 DR SMART: SM00408: IGC2_14.
 DR SMART: SM00281: Lamb_3.
 DR SMART: SM00282: Lamb_3.
 DR SMART: SM00192: LDLa_4.
 DR SMART: SM00200: SEA_1.
 DR PROSITE: PS00022: EGF_1; 8.
 DR PROSITE: PS01186: EGF_2; 5.
 DR PROSITE: PS01245: LAMININ_TYR_EGF_11.
 DR PROSITE: PS01209: LDLA_E_4.
 DR PROSITE: PS00068: LDLA_2_4.
 DR PROSITE: PS00024: SEA_1.
 KW Signal: Basement membrane; Proteoglycan; Recept; Glycoprotein;
 KW Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
 KW Extracellular matrix; EGF-like domain.
 FT CHAIN 1 21
 FT SIGNAL 1 21
 FT 22 3707
 FT 1 80 194
 FT DOMAIN 195 234 LDL-RECEPTOR CLASS A 1.
 FT 241 419 LDL RECEPTOR CLASS A 2.
 FT DOMAIN 420 459 LDL-RECEPTOR CLASS A 3.

DR Pfam: PF00054; laminin_G: 3.
 DR Pfam: PF00057; ldl_recept_a: 4.
 DR Pfam: PF01390; SEA: 1.
 DR PRINTS: PK00010; EGF_MCOH.
 DR ProDom: PD003031; laminin_B: 3.
 DR SMART: SM00180; EGF_Lam: 6.
 DR SMART: SM00001; EGF_Like: 8.
 DR SMART: SM00406; lcc2: 22.
 DR SMART: SM00281; Lam: 3.
 DR SMART: SM00282; lamg: 3.
 DR SMART: SM00192; ldlra_4: 4.
 DR SMART: SM00200; SEA: 1.
 DR PROSITE: PS00022; EGF_1: 9.
 DR PROSITE: PS01186; EGF_2: 5.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF_11.
 DR PROSITE: PS01209; ldlra_1: 4.
 DR PROSITE: PS50068; ldlra_2: 4.
 DR PROSITE: PS50024; SEA: 1.
 DR Signal: Basement membrane; Proteoglycan; heparan; glycoprotein;
 KM Heparan sulfate; laminin EGF-like domain; immunoglobulin domain;
 KM Extracellular matrix; EGF-like domain.
 PF CHAIN 1 21
 PF SIGNAL 22 4393
 PF DOMAIN 80 194 LDL-RECEPTOR CLASS A 1.
 PF DOMAIN 197 246 LDL-RECEPTOR CLASS A 2.
 PF DOMAIN 283 323 LDL-RECEPTOR CLASS A 3.
 PF DOMAIN 323 361 LDL-RECEPTOR CLASS A 4.
 PF DOMAIN 366 405 IG-LIKE C2-TYPE DOMAIN 1.
 PF DOMAIN 405 506 LAMININ EGF-LIKE 1 (N-TERMINAL).
 PF DOMAIN 533 532 LAMININ EGF-LIKE 1 (N-TERMINAL).
 PF DOMAIN 533 742 LAMININ EGF-LIKE 1 (N-TERMINAL).
 PF DOMAIN 742 765 LAMININ EGF-LIKE 1 (C-TERMINAL).
 PF DOMAIN 765 815 LAMININ EGF-LIKE 2.
 PF DOMAIN 815 873 LAMININ EGF-LIKE 3.
 PF DOMAIN 873 925 LAMININ EGF-LIKE 4 (INCOMPLETE).
 PF DOMAIN 925 935 LAMININ EGF-LIKE 5 (N-TERMINAL).
 PF DOMAIN 935 935 LAMININ EGF-LIKE 5 (N-TERMINAL).
 PF DOMAIN 935 1127 LAMININ EGF-LIKE 5 (C-TERMINAL).
 PF DOMAIN 1127 1160 LAMININ EGF-LIKE 5 (C-TERMINAL).
 PF DOMAIN 1160 1210 LAMININ EGF-LIKE 6.
 PF DOMAIN 1210 1267 LAMININ EGF-LIKE 7.
 PF DOMAIN 1267 1326 LAMININ EGF-LIKE 8.
 PF DOMAIN 1326 1336 LAMININ EGF-LIKE 9 (N-TERMINAL).
 PF DOMAIN 1336 1531 LAMININ EGF-LIKE 9 (N-TERMINAL).
 PF DOMAIN 1531 1531 LAMININ EGF-LIKE 9 (N-TERMINAL).
 PF DOMAIN 1531 1564 LAMININ EGF-LIKE 9 (C-TERMINAL).
 PF DOMAIN 1564 1614 LAMININ EGF-LIKE 10.
 PF DOMAIN 1614 1672 LAMININ EGF-LIKE 11.
 PF DOMAIN 1672 1773 IG-LIKE C2-TYPE DOMAIN 2.
 PF DOMAIN 1773 1867 IG-LIKE C2-TYPE DOMAIN 3.
 PF DOMAIN 1867 1957 IG-LIKE C2-TYPE DOMAIN 4.
 PF DOMAIN 1957 2053 IG-LIKE C2-TYPE DOMAIN 5.
 PF DOMAIN 2053 2153 IG-LIKE C2-TYPE DOMAIN 6.
 PF DOMAIN 2153 2246 IG-LIKE C2-TYPE DOMAIN 7.
 PF DOMAIN 2246 2342 IG-LIKE C2-TYPE DOMAIN 8.
 PF DOMAIN 2342 2438 IG-LIKE C2-TYPE DOMAIN 9.
 PF DOMAIN 2438 2535 IG-LIKE C2-TYPE DOMAIN 10.
 PF DOMAIN 2535 2631 IG-LIKE C2-TYPE DOMAIN 11.
 PF DOMAIN 2631 2728 IG-LIKE C2-TYPE DOMAIN 12.
 PF DOMAIN 2728 2828 IG-LIKE C2-TYPE DOMAIN 13.
 PF DOMAIN 2828 2926 IG-LIKE C2-TYPE DOMAIN 14.
 PF DOMAIN 2926 3023 IG-LIKE C2-TYPE DOMAIN 15.
 PF DOMAIN 3023 3114 IG-LIKE C2-TYPE DOMAIN 16.
 PF DOMAIN 3114 3213 IG-LIKE C2-TYPE DOMAIN 17.
 PF DOMAIN 3213 3300 IG-LIKE C2-TYPE DOMAIN 18.
 PF DOMAIN 3300 3401 IG-LIKE C2-TYPE DOMAIN 19.
 PF DOMAIN 3401 3490 IG-LIKE C2-TYPE DOMAIN 20.
 PF DOMAIN 3490 3576 IG-LIKE C2-TYPE DOMAIN 21.
 PF DOMAIN 3576 3671 IG-LIKE C2-TYPE DOMAIN 22.
 PF DOMAIN 3671 3701 LAMININ G-LIKE 1 (GLOBULAR DOMAIN V A).
 PF DOMAIN 3701 3843 EGF-LIKE 1.
 PF DOMAIN 3843 3883 EGF-LIKE 2.
 PF DOMAIN 3883 3924 EGF-LIKE 3.
 PF DOMAIN 3924 4104 LAMININ G-LIKE 2 (GLOBULAR DOMAIN V B).
 PF DOMAIN 4104 4143 EGF-LIKE 3.

PF DOMAIN 4145 4178 EGF-LIKE 4.
 PF DOMAIN 4243 4391 LAMININ G-LIKE 3 (GLOBULAR DOMAIN V C).
 PF SITE 65 67 HEPARAN SULFATE (POTENTIAL).
 PF SITE 71 73 HEPARAN SULFATE (POTENTIAL).
 PF SITE 76 78 HEPARAN SULFATE (POTENTIAL).
 PF SITE 4151 4153 MEDIANES MOTOR NEURON ATTACHMENT (POTENTIAL).
 PF SITE 4301 4303 MEDIANES MOTOR NEURON ATTACHMENT (POTENTIAL).
 PF SITE 199 212 HY SIMILARITY.
 PF DI-SOLID 206 225 HY SIMILARITY.
 PF DI-SOLID 219 234 HY SIMILARITY.
 PF DI-SOLID 285 297 HY SIMILARITY.
 PF DI-SOLID 292 310 HY SIMILARITY.
 PF DI-SOLID 404 419 HY SIMILARITY.
 PF DI-SOLID 425 337 HY SIMILARITY.
 PF DI-SOLID 332 350 HY SIMILARITY.
 PF DI-SOLID 344 359 HY SIMILARITY.
 PF DI-SOLID 368 381 HY SIMILARITY.
 PF DI-SOLID 375 394 HY SIMILARITY.
 Query Match 5,882 Score: 102.59 DB 1: Length 4393
 Local Local Similarity: 24,176 Ident: 6,662
 Matches: 84; Conserved: 69; Mismatches: 140; Indels: 85; Gaps: 14;
 QY 48 TVVGTNKLITGLDSCNYEALITLTKNDI-----VRRRLK-----GVYKRE 92
 DB 2257 TVAEQQTLLISCV-ACQIAQVTVYRKGSIDVARIQKSLYTPQASIMQGVYCA 2415
 QY 93 YGGLTSSGASLSTPCTGCA-----P-I-HPTSSKRYFNITTR----- 129
 DB 2316 SNGMCA--TVVTVTGCAALAVYAGSTVTRLEPSSQVAEQQTLLDNCVINGSHQA 2374
 QY 130 ---HKMVALPVMASSTPSPSKTLKLCIDVPMQVFDITCQILCKRAVDASFLKRWG 186
 DB 2374 VTMIRKGSLSLVIRIQTICS---LIRIYQASIVASGECVGLSSVPLASVLTTHPA 2429
 QY 187 TKTP-----PFSWRVLGDVLEGLSHIIRLQNTITDILVTDN-HIVV-----AK 231
 DB 2410 GSVVALGVTVYVKI-----ESSSSQVAEQQTLLDNCVIAQVIAVIMIRKGSIPAK 2482
 QY 2342 SLKSGSLFKI-----YSLTKRLQSSMSSENOTMLSLFPHIGTSYGRGIRVLPES 281
 DB 2483 HQVIGSKRLDQVTVASGECVCAVVGSSGTQELASVLTVOQLSGSHGCAVTVRIIES 2542
 QY 282 NSDVQDLKDLSEANIYANQHSQVIGSETPDSPTN---GVSLRPP 324
 DB 2543 SSASLAWGHITDNCVIAVSQVHTITWYRKGSLSIRHQVGSRLRTP 2590
 RESULT 8
 YH47_METUA
 ID YH47_METUA STANDARD; PRT: 462 AA.
 AC OS8547;
 DT 01-NOV-1997 (Ref. 35, Cited)
 DT 01-NOV-1997 (Ref. 35, Last sequence update)
 DT 20-AUG-2001 (Ref. 40, Last annotation update)
 DE HYPOTHEICAL PROTEIN M1147.
 GN M1147.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
 OC Methanococcus.
 OX NCBI_TaxID:2190;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43047;
 RX MEDLINE:96337999; PubMed:868087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kervajog A.R., Dougherty B.A., Tomb J.-F., Adams M.D.,
 RA Overberg K., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Gooden N.S.M., Weidman J.F., Fuhlmann J.L., Nguyen D.,
 RA Bitterback T.K., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,

Query Match	6.1k	Score 109	Dr 1	Length 569
11	SEQUENCE FROM N.A.			
12	MEDLINE:92218664; PubMed-13136466;			
13	Wang P.L., O'Farrell S., Clayberger C., Kravsky A.M.:			
14	"Identification and molecular cloning of tactile, A novel human T			
15	cell activation antigen that is a member of the Ig gene			
16	superfamily".			
17	J. Immunol. 148:2600-2608(1992).			
18	-1- FUNCTION: MAY BE INVOLVED IN ADHESIVE INTERACTIONS OF ACTIVATED T			
19	CELLS.			
20	AND NK CELLS DURING THE LATE PHASE OF THE IMMUNE RESPONSE. MAY			
21	FUNCTION AT A TIME AFTER T AND NK CELLS HAVE GENERATED THE			
22	ENVIRONMENT USING INTEGRINS AND SELECTINS, WHEN THEY ARE ACTIVELY			
23	ENGAGING DISEASED CELLS AND MOVING WITHIN AREAS OF INFLAMMATION.			
24	-1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.			
25	-1- SUBCELLULAR LOCATION: TYPE 1 MEMBRANE PROTEIN.			
26	-1- TISSUE SPECIFICITY: EXPRESSED ON NORMAL T CELL LINES AND CLONES,			
27	AND SOME TRANSFORMED T CELLS, BUT NO OTHER CULTURED CELL LINES.			
28	THEYED. IT IS EXPRESSED AT VERY LOW LEVELS ON ACTIVATED T CELLS.			
29	-1- DEVELOPMENTAL STAGE: EXPRESSED AT LOW LEVELS ON PERIPHERAL T CELLS			
30	AND IS STROKING UP-REGULATED AFTER ACTIVATION, PEAKING 6 TO 9 DAYS			
31	AFTER THE ACTIVATING STIMULUS.			
32	-1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.			
33	-1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.			
34	-1- DATABASE: NAME-PROX; NOT-CD guide CDR6 entry.			
35	WWW-HTTP://WWW.NCBI.NLM.NIH.GOV/PROX/CD/CD96.HTM".			
36	-----			
37	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
38	between the Swiss Institute of Bioinformatics and the EMBL collaboration.			
39	CC The European Bioinformatics Institute. There are no restrictions on its			
40	CC use by non-profit institutions as long as its content is in no way			
41	CC modified and this statement is not removed. Usage by and for commercial			
42	CC entities requires a license agreement (see http://www.isb-sib.ch/announcement/			
43	CC or send an email to license@isb-sib.ch).			
44	CC			
45	DR EMBL: M88282; AAA36662.1;			
46	DR PIR: A46462; A46462.			
47	DR InterPro: IPR003599; 19.			
48	DR InterPro: IPR003006; 19_MHC.			
49	DR Pfam: PF00047; 19; 1.			
50	DR SMART: SM00409; 19; 1.			
51	DR Immunoglobulin domain; Glycoprotein; Transmembrane; Signal.			
52	KW SIGNAL			
53	FT CHAIN	1	21	POTENTIAL.
54	FT DOMAIN	22	569	T-CELL SURFACE PROTEIN TACTILE.
55	FT TRANSMEM	22	501	EXTRACELLULAR (POTENTIAL).
56	FT DOMAIN	502	525	POTENTIAL.
57	FT DOMAIN	526	569	CYTOPLASMIC (POTENTIAL).
58	FT DOMAIN	534	486	PRO-STER/TIR-RICH.
59	FT DOMAIN	538	547	PRO-RICH.
60	FT DOMAIN	38	125	1G-LIKE V-TYPE DOMAIN 1.
61	FT DOMAIN	156	238	1G-LIKE V-TYPE DOMAIN 2.
62	FT DOMAIN	267	346	1G-LIKE C2-TYPE DOMAIN.
63	FT DISULFID	45	118	PROBABLE.
64	FT DISULFID	163	231	PROBABLE.
65	FT DISULFID	274	339	PROBABLE.
66	FT CARBOHYD	42	42	N-LINKED (GLCNAC. . .) (POTENTIAL).
67	FT CARBOHYD	97	97	N-LINKED (GLCNAC. . .) (POTENTIAL).
68	FT CARBOHYD	107	107	N-LINKED (GLCNAC. . .) (POTENTIAL).
69	FT CARBOHYD	148	148	N-LINKED (GLCNAC. . .) (POTENTIAL).
70	FT CARBOHYD	156	156	N-LINKED (GLCNAC. . .) (POTENTIAL).
71	FT CARBOHYD	166	166	N-LINKED (GLCNAC. . .) (POTENTIAL).
72	FT CARBOHYD	184	184	N-LINKED (GLCNAC. . .) (POTENTIAL).
73	FT CARBOHYD	199	199	N-LINKED (GLCNAC. . .) (POTENTIAL).
74	FT CARBOHYD	261	261	N-LINKED (GLCNAC. . .) (POTENTIAL).
75	FT CARBOHYD	262	262	N-LINKED (GLCNAC. . .) (POTENTIAL).
76	FT CARBOHYD	284	284	N-LINKED (GLCNAC. . .) (POTENTIAL).
77	FT CARBOHYD	334	334	N-LINKED (GLCNAC. . .) (POTENTIAL).
78	FT CARBOHYD	352	352	N-LINKED (GLCNAC. . .) (POTENTIAL).
79	FT CARBOHYD	419	419	N-LINKED (GLCNAC. . .) (POTENTIAL).
80	FT CARBOHYD	481	481	N-LINKED (GLCNAC. . .) (POTENTIAL).
81	SEQUENCE	569 AA:	63887 MM:	DP2F3B8E350F3BF2 CRG04:

```

Host Loca Similarity 20.7%: Pred. no. 0.1:
Matches 78; ConservedLive 71; Mismatches 145; Indels 82; Gaps 20.

OY 16 GJRGCTGVVNYGVVKKPKVYLSKGTNY-----SVYTVV--QTNVELTCLT----- 61
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 55 QMWSRYTNKIDELAVNYGV-----GVYCAVGGVCSLWTFETEPENGSKWFLHLKMSKST 110
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 62 -PGCNVLAFLVTKNDIVAFHKLGVYKRELVQVYSSGQVASTPFRCHLAAIIPRYS 120
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 111 NVNARVTKRLVLAFLPGLTRFVHLVGVTVVAVRMSKSTITTEELNGLTILPFGNSS 166
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 121 KY-----FNPFPDHLKRVALLVWASVTHMSPSWTLAK-----LCPVGGKGVPT-- 164
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 170 KLSSELYVAVSWVDCNGTETL--ISGNILSNSTLAKDVKLCTYTRKDLHSLVGFIDCKR 227
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 164 -LTCGL-LKRAVWASSTLAKWKKVETFPFSKRWLIG-----PLVLRDQ-----LSLILRD 214
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 228 KPSGHLKVCNKLKSSPFRKVPKKEFL--VIVENNSTVIVKRGKTLKRVPR 282
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 214 QNCT--LD--LNAVLDN--VHVASRLKVG--STLQVSLPRLKLS--MNSPQNTMLSLAT 263
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 283 ANTWVLRKSLIDDEKGLVTFNNEKRCNGCELEKSVLTQVHSNKPAGSDMLTQWCM-- 340
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 264 ILACGCTSYCGKIKVLPESNSVDVQIKRDLPSANLTPANGSDVLCSEFPHSPKVCVSLAP 323
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 341 -----ALSTVQNKV-----WNLSSEKTPPLATSELSSTDPPLSVTSESLDQPL 384
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 324 PCWSEVARSRLVAST 339
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 385 SPASSVAPKATVTS 400

RESULT 6
EXOC_HPV4 STANDARD: PCT: 560.0A.
1D EXOC_HPV4 AC P04522:
DT 13-AUG-1987 (Ref. 05, Cited)
DT 01-JAN-1990 (Ref. 13, Last sequence update)
DT 20-AUG-2001 (Ref. 40, Last annotation update)
DE EXONUCLEASE SUBUNIT 2 (EC 3.1.11.1) (PROTEIN HPV4).
GN 46.
OS Bacteriophage T4.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae;
OC T4-like phages.
OX NCBI_TaxID:10665;
RN 111
RP SEQUENCE FROM N.A.
RX MEDLINE=85257446; PubMed 4018026;
RA Gram B., Roeger W.;
RT "Genes 55, alpha qL 47 and 46 of bacteriophage T4: The genomic
    organization as deduced by sequence analysis.";
RL EMDJ. 4:257-264(1985).
RN 121
RP SEQUENCE FROM N.A.
RA Kuter E., Arisaka F., Kunisawa T., Tsugita A., Mosig G.,
    Meschaninov V., Roeger W., Stidham T., Thomas E.;
RT "Bacteriophage T4 genome analysis.";
RL Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.
RN 131
RP SEQUENCE OF 478-560 FROM N.A.
RA Hsu T., Katam J.;
RL Submitted (APR-1987) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: EXONUCLEASE INVOLVED IN PHAGE DNA RECOMBINATION,
    REPLICATION AND REPAIR.
CC -1- SUBUNIT: CONSISTS OF TWO SUBUNITS: GP47 AND GP46.
CC -1- SIMILARITY: STRONG TO TWO PROTEIN D13 AND TO YEAST RAD52.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
    between the Swiss Institute of Bioinformatics and the EMBL outstation -
    the European Bioinformatics Institute. There are no restrictions on its
    use by non-profit institutions as long as its content is in no way
    modified and this statement is not removed. Usage by and for commercial
    entities requires a license agreement (See http://www.isb-sib.ch/announc/)

```

RT *Euplotes crassus has genes encoding telomere-binding proteins and
 RL telomere-binding protein homologs *;
 CC Nucleic Acids Res. 20:6621-6629(1992).
 CC -|- FUNCTION: MAY BIND TELOMERIC TAG4 SEQUENCES.
 CC -|- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -|- SIMILARITY: BELONGS TO THE TELOMERE-BINDING PROTEIN ALPHA SUBUNIT
 CC FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M96819; AAA29128.1; -
 DR InterPro: IPR003415; Telo_bind_alpha.
 DR Pfam: PF02307; Telo_bind_alpha; 1.
 KW DNA-binding; Nuclear protein; Telomere; Multigene family.
 SQ SEQUENCE 460 AA: 53360 MW: EDIC141385AUB5PB CRC64;
 Query Match 8.0%; Score 142.5; DB 1; Length 460;
 Best Local Similarity 22.7%; Pred. No. 9.9e-05;
 Matches 73; Conservative 45; Mismatches 118; Indels 85; Gaps 13;
 QY 9 YIYTPNLQ--KGGTIVNVYGVVKKFPKPYLSTGDTYCVSVTVIVDT-NVK-----LTC 59
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 QY 19 YEYTEIGSTEENEASINFYAVVADACFPYKVKDEKYMVYLKVIDTTINVRKXGDNPATV 78
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 QY 60 LLFSCNYEALPIIYKNGDIVRHLK IOVYKKEGTTI-----SSGFASLTFFGTLAGPI 114
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 QY 79 ALOSRRFEDLPICQGDLCIRHRAEYN-YKDDQHYFKLNSSYSWALFSADREVAPEV 137
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 QY 115 IP-----RTSSKYFNFTEDIKMVEALRWASTHSPSWTLKLCIDVQPMQVFD 163
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 QY 138 LKDEGDDFTYRSYAVSGKYQNFDTQDOKLLNKNTRANKSYAKN-----DV----- 183
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 QY 164 LTCCLLKAEVNGASFLKVDGTRTPPPSWRVLIQDLVLKGDLSH LIRLQNLTDILVY 223
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 QY 184 -----LIDENYTPLSQARQEGDFNVVGVKVTQIVHRYVYTSDLRVK 224
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 QY 224 DN-----HVVHVARSLKVGSEFIRY-SLIJTKLOSMSSENOTMISLEPHLRGTSYGRGIRV 277
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 QY 225 DTSKATWLTYSRR-----KPRRLTEGVITIKRSVNIHSE-----TERERCELE 268
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 QY 278 LPESN-----SDVDQLKKDLKS 294
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 QY 269 APHSNIMTFVDFSLAKSLDS 289
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 RESULT 4
 TEHA_STYMY STANDARD: PRT: 493 AA.
 AC P29550;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE TELOMERE-BINDING PROTEIN ALPHA SUBUNIT.
 GN STY56V AND STY56L.
 OS Styl mychia mytilis.
 OC Eukaryota; Alveolata; Ciliophora; hypotrichs; Stichotrichida;
 OC Oxytrichidae; Stylonychia.
 OX NCBI_Taxid=5952;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92051290; PubMed=1945829;
 RA Fang G., Cech T.R.;
 RT "Molecular cloning of telomere-binding protein genes from Stylonychia
 mytilis.";
 RL Nucleic Acids Res. 19:5515-5518(1991).

CC -|- FUNCTION: MAY FUNCTION AS PROTECTIVE CAPPING OF THE SINGLE-
 CC STRANDED TELOMERIC OVERHANG. MAY ALSO PARTICIPATE IN TELOMERE
 CC LENGTH REGULATION DURING DNA REPLICATION.
 CC -|- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
 CC -|- SUBCELLULAR LOCATION: NUCLEAR.
 CC -|- MISCELLANEOUS: THE SEQUENCE OF STY56V IS SHOWN. STY56L DIFFERS IN
 CC ONLY ONE POSITION.
 CC -|- SIMILARITY: BELONGS TO THE TELOMERE-BINDING PROTEIN ALPHA SUBUNIT
 CC FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X61749; CAA43888.1; -
 DR PIR: S22581; S22581.
 DR InterPro: IPR003415; Telo_bind_alpha.
 DR Pfam: PF02307; Telo_bind_alpha; 1.
 KW DNA-binding; Nuclear protein; Telomere; Multigene family.
 FT VARIANT 429 429 V->I (IN STY56L).
 SQ SEQUENCE 493 AA: 56097 MW: B49002984363CSAA CRC64;
 Query Match 7.6%; Score 135; DB 1; Length 493;
 Best Local Similarity 21.1%; Pred. No. 0.00048;
 Matches 79; Conservative 57; Mismatches 148; Indels 90; Gaps 15;
 QY 6 ATNVIYTPLNQ--LKGTTIVNVYGVVKKFPKPYLSTGDTYCVSVTVIVD----- 51
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 30 STRYKVELAKASITSAEAOHFYGVVVDATFPYKTNQERYICSLKVVDPDSLKLSOKGTG 89
 QY 52 QTNVKLTCLFSCNYEALPIIYKNGDIVRHLK IOVYKKEGTTI-----GLTSSG 100
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 90 DASDYATLVLYAKRFEDLPICQGDLCIRHRAEYN-YKDDQHYFKLNSSYSWALFSADREVAPEV 149
 QY 101 FASLTFFGTLAGPI (PRT)-----SSKYFNFTEDIKMVEALRWASTHIM-----SPSWT 149
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 150 KKSALQETIGGEPASDLTPFAFSGKNYTFEKSSEALLONIRKWAQVFOQYVNISSDMET 209
 QY 150 LKLCIDVQPMQVFDLTCCLLKAEVNGASFLKVDGTRTPPPSWRVLIQDLVLKGDLSH 209
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 210 PLKAAQAKGSD-FVYVAKLQAVFELDEYTNELKLDQSGQVF-----VTLAKLKFPH 261
 QY 210 IH-----RIQNLTDID-----LLVYDNIVHVARSLKVGSEFIRYSLIHTKLOSMSSENOT 257
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 262 LRAGEVVRIRSATYDETSTOKKVVLLLSHYNSNIVTFVSASKLAK--EVKAKVTUDKSVKKA 319
 QY 258 MLSLEPHL-----HGCTSYGRGIRVLPF-----SNSVDQLKKDLKSANLTA 299
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 320 ALKQDVSLSAVVLTVDKKHAG-----LPTHSIQDLFINADTD---KEISSKDTFR 367
 QY 300 NOHSDVMCOSEPPD 413
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 368 TQF--YITRVPVD 479
 RESULT 5
 TACT_HUMAN STANDARD: PRT: 569 AA.
 AC P40200;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE T-CELL SURFACE PROTEIN TACTILE PRECURSOR (CD96 ANTIGEN).
 GN CD96.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_Taxid=9606;


```

QY 164 -LTCOL-LGRKAVDCAFLKAWDGTPTTPSWKVLIO-----DAVLEGI-----LSHHLKL 273
DB 228 KPSCHIKKPGFNKTLKSTSTVVKVFAKPEIP-----VJVENNSTVJVEKGFPTCLKAVIPK 282
QY 214 ONLT--ID-ILVYDNH--VHVARSLEKVC--SEPLKIVSLTKLQK--MNSSENQMLSLPEP 263
DB 283 ANITWPTLQSGSLHDKKCIYITNEERKGGKDPLEKSVLTVHNSKRPVQSPNLTTCWCH-- 340
QY 264 HILGCTSTGKCKRVLDPESNSVDQI-KKDLKSNLTVANQSHVTCQSEPDSDSPNCVSLAP 328
DB 341 -----ALSPVQCNKV-----WNISSEKTPPLKGSLSISPDPLSVTESSTLDPDP 384
QY 324 JKMSVARSRLIAAST 339
DB 365 SPASSVSPAKYATVSS 400

RESULT 8
T22006
hypothetical protein F39H2.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #cexl_change 15-Oct-1999
C:Accession: T22006
K:Korshaw, J.
submitted to the EMBL data library, October 1996
A:Reference number: Z19501
A:Accession: T22006
A:Status: preliminary; translated from GH/EMBL/DBM
A:Molecule type: DNA
A:Residues: 1-587 <WTL>
A:Cross-references: EMBL:Z81080; P1DN:CAH03086.1; GSPDB:GN00019; CESP:F39H2.1
A:Experimental source: clone F39H2
C:Genetics:
A:Gene: CESP:F39H2.1
A:Map position: 1
A:Introns: 42/3; 104/2; 160/3; 273/2; 317/2; 429/2; 502/2

Query Match 6.1%; Score 109; DB 2; Length 587;
Host Local Similarity 22.9%; Pred. No. 0.29;
Matches 51; Conservative 24; Mismatches 60; Indels 88; Gaps 10;

QY 122 YFNFTEDHKWE-----ALRWASTHMSI-SWTLKLCIVQPMQYEDP-----COL-- 168
DB 71 YFDVLAQVHSAVETKNGSWTLRWAKQKCPKESKER-----KENMLPVTENTPKCYIVP 126
QY 169 ----LGRN-EVDGASPLKAVDGTPTTPSWKVLIODLVLKGLSHIHLQNLITIDLIV 223
DB 127 PDPRIKAIVESGEFL-----IETIDV 149

QY 224 DNHHVARSLEKVSFLKIVSLH-----TKLQSNSENOFMLSLFPHDICTSYGCGIRV 277
DB 150 DDHRIQIEKLNSGDFVALQNVHASVGLTDMQVLG-----GGQAVNRKISK 196

QY 278 LPES--NSDVQLLKDLKESA-----NLTVANQSHVTCQSE 310
DB 197 VPDPFRNEARFQIKKRVESVLEAVTDCINPLEFOCKENNVQSE 239

RESULT 9
T00342
hypothetical protein KIA00580 - human (fragment)
C:Species: Homo sapiens (hmn)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #cexl_change 21-Jul-2000
C:Accession: T00342
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
DNA Res. 5, 31-39, 1998
A>Title: Prediction of the coding sequences of unidentified human genes. IX. The complete
A:Reference number: Z14086; M01D:98290545
A:Accession: T00342
A:Status: preliminary; translated from GH/EMBL/DBM
A:Molecule type: mRNA
A:Residues: 11044 <NAG>

```

```

A:Cross-references: EMBL:AB01152; NID:q3043683; P1DN:BAZ5506.1; P1D:q3043684
A:Experimental source: brain; clone HJ0601
C:Genetics:
A:Gene: KIA00580

Query Match 5.8%; Score 104; DB 2; Length 1044;
Host Local Similarity 24.0%; Pred. No. 2.2;
Matches 70; Conservative 39; Mismatches 119; Indels 64; Gaps 15;

QY 49 LVQDNVKTFCIL-----QSGNEALDILYKNCIDVRIHKLQVKKRKTQCTSSGASL 105
DB 245 LDEETN-KKRCVLAKGCLSYTENKSTTPNCTTQINVCIALHKEDEYANTCPTP-- 410
QY 106 PEGTLAGATTPRTSKRYNPPTTEHKEVEALRWVA-----STHMSWTLAKLCIVQPMQY 162
DB 311 PE-----IYLPSEKVFLECATFS--GQGRKTEALIANHVFILVA-EM/TPAD-----Y 455
QY 163 DLTCOLLKRAVAKASPLKAV-----DKTPTTPSWKVLIODLVLKGLSHIHLQNLITD 219
DB 346 DLTCOLLKRAVAKASPLKAV-----DKTPTTPSWKVLIODLVLKGLSHIHLQNLITD 219
QY 220 ILVYDNH-VHVARSLEKVSFLKIVSLTKL-----QSMNSENO 257
DB 410 TMYGNGEKELVILLVERKRTYTHG-HPRLPVTWVITALEKAGCTKRNALQVQGLSKNIV 468
QY 258 MSLSEPHLHGTSTYGCRLVLPESNSD-----VQIKKDLKSNLTVANQ 402
DB 469 P1TVNSCLAVTQVGLCKKLYQNRKQDLHSLSEPKKIAVSFKLRGRH 520

RESULT 10
MCHPX6
exonuclease 4b (EC 4.1.11-) chain 2 - phage T4
C:Species: phage T4
A:Note: host Escherichia coli
C:Date: 17-Mar-1987 #sequence_revision 21-Nov-1997 #cexl_change 19-Jan-2001
A:Accession: A04298; T10162
K:Graham, H.; Koenig, W.
EMBO J. 4, 257-264, 1985
A>Title: Genes 55, alpha-61, 47 and 46 of bacteriophage T4: the genome organization
A:Reference number: A91016; M01D:85267446
A:Accession: A04298
A:Molecule type: DNA
A:Residues: 1-560 <GRA>
A:Cross-references: GR:X07804; NID:q15229; P1DN:CAZ5945.1; P1D:q577852
A:Note: the authors translated the initiation codon GUG for residue 1 as Val
C:Genetics:
A:Gene: 46
A:Start codon: GUG
C:Superfamily: phage T4 exonuclease 4b
C:Keywords: exonuclease; hydrolase; nucleotide binding; P-loop
P:36-43/Region: nucleotide-binding motif A (P-loop)

Query Match 5.8%; Score 102.5; DB 1; Length 560;
Host Local Similarity 21.0%; Pred. No. 0.97;
Matches 63; Conservative 46; Mismatches 98; Indels 93; Gaps 12;

QY 65 NYEALPILYKNCIDVRIHKLQVKKRKTQCTSSGASL 117
DB 3 NFKLNVKRYKNIINSVQNCIDLDLVKQKTLTGRMKCKSTPLMLATP-GLFGKPRHY 61
QY 118 TSKRYNPPTTEHKEVEALRWVA-----STHMSWTLAKLCIVQPMQY-----F 162
DB 62 KKCOLINSTNKKELIVE--LW-----MEYDEKKYIKRGOKPNVF 99
QY 163 DLTC-----QLAKRAVAKASPLKAVMKTTR--PPE-----PSMRV 196
DB 100 EITVNGTKLNSASSKDPQAEFPOLLG--MSTASPKQIVLGTACTGTPMGLSTPAARK 156
QY 197 LIODLVLECDLSHIIHLQNLITDILVYDNHIVARSLEKVSFLKIVSLTKLQ--SMNSEN 255

```

Db 138 IKDEGDDFTYRSAYSGKYNFTDOKLLKNTRAWKSYFAKN-----DV----- 183
 QY 164 LTCQLLCKAEVDGASFLKLVWDGTRTPPSWRVLQDLVLEGLDSHHLRLQLNLTIDILVY 223
 Db 184 -----IIDEMVTPUSQARQEGDFNVGKVQIVRHRYDYYTSDLAVK 224
 QY 224 DN-----HVRVARSKVGSLRIY-SLHTKQSMNSNQTMLSLEFHLHGTSYGRGIRV 277
 Db 225 DTSKATWFLTVSR-----KFPRLYEGVLIKIRSVNIDSE-----TERERCLLEL 268
 QY 278 LPESN-----SDVDLKKDLES 294
 Db 269 APHSNIMTFPFSRLAKSLDS 289

RESULT 5
 S22581
 telomere-binding protein alpha chain - Stylonychia mytilus
 C:Species: Stylonychia mytilus
 C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 07-Dec-1999
 C:Accession: S22581
 R:Fanig, G.; Cech, T.R.
 Nucleic Acids Res. 19, 5515-5518, 1991
 A:Title: Molecular cloning of a telomere-binding protein genes from Stylonychia mytilus.
 A:Reference number: S22580; MUID:92051290
 A:Accession: S22581
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-493 <FAN>
 A:Cross-references: EMBL:X61749; NID:q10194; PTD:q578501
 C:Genetics:
 A:Genetic code: SCS5
 A:Introns: 32/3
 C:Keywords: DNA binding; nucleus

Query Match 7.64; Score 135; DB 2; Length 493;
 Best Local Similarity 21.18; Pred. No. 0.0014;
 Matches 79; Conservative 57; Mismatches 148; Indels 90; Gaps 15;

QY 6 ATNYITPLNQ--LKGCTIVNVYGVVKKFPKPPYLSKGTGYGVVTIVD----- 51
 Db 30 STRYKRVVNLKASLTSAEAQHYGVV IDATPEYKTNQERYTCSLKVVDPSPSLKSKQCTG 89
 QY 52 OTNVKLTCLLSCNVEALPIYKNGDIVRFIRLKLQVYKHTQ-----GTTSSG 100
 Db 90 DASDVTILVYAKRFEDLPIIIRIGDITVRHATRLRYNGQRFNANVFYNSSWALFSTD 149
 QY 101 FASLTFEGLTGAIPRT-----SSKYENFTTEDHKMVEALRVWASTIUM-----SPSWT 149
 Db 150 KKSALQELGGOPASDITPPAFSCKNYTPFKSEANLIONIRKWAQVFOQYNYVSSDMPT 209
 QY 150 LKRLGIDVQPMQYFDLTQLLAKAEVDGASFLKLVWDGTRTPPSWRVLQDLVLEGLDSH 209
 Db 210 PLNKAAQAKGD-FDQVAKILQVFELEDYTNELKLDQSGQVF-----YTLALKLKFPH 261
 QY 210 HI-----RIQNIITD-----ILVYDNIHIVARSLKVGSPFLRYSLHTKQSMNSNQ 257
 Db 262 LRAGRVVLRATYDTSSTOKKVLULLSHYSNTVTFVSASKLAK-EVKAKVTIQRVSKA 319
 QY 258 MSLSEFHL-----HGTSYGRGIRVLP-----SNSVDVQLKKDLESANLTA 299
 Db 320 ALKODVLSLAVLTFVDKKHAG-----LPTHSLODLFHADTD---KELSSKUTFR 367
 QY 300 NOHSDVICOSEPD 313
 Db 368 TQF---YITRVEPD 379

RESULT 6
 t22839
 hypothetical protein F57C2.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: t22839
 R:Baynes, C.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: t219624
 A:Accession: t22839
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-251 <WIL>
 A:Cross-references: EMBL:Z83110; PIDN:CAB05525.1; GSPDB:GN00020; CESP:F57C2.3
 A:Experimental source: clone F57C2
 C:Genetics:
 A:Gene: CESP:F57C2.3
 A:Map position: 2
 A:Introns: 115/3; 198/2

Query Match 7.34; Score 130; DB 2; Length 251;
 Best Local Similarity 25.74; Pred. No. 0.0014;
 Matches 52; Conservative 10; Mismatches 66; Indels 54; Gaps 10;

QY 160 QYEDLTCLLCKAEVDGASFLKLVWDG-----PKTFPPSMRV----- 196
 Db 25 RYEDVLAQVHVVVETINGAQLRWVRGRKFGPGASAEHRKRRRLHFHVTDQSFKRYIVPPNP 84
 QY 197 LIODLVLEGLDSHHLRLQNLITIDILVYDNIHIVARSLKVGSPFLRYSLHTKQSMNSNQ 256
 Db 85 RICKATEENG-----KELLIFIDYDHRGLKWLNSGDFVATONVHAA-----STRQ 132
 QY 257 TMLSLEFHLH-GGTSYGRGIRVLP-----SNSVDVQLKKDLES-----NLTAHQHSDV 305
 Db 133 TEMQV---LHGGSASYQGITTPVPDPEHEAFQNEFKKKEAVLETVAVDENFTFOQPEE 189
 QY 306 ICOS-----EP-DDSPNGVSLR 322
 Db 190 VAENHVDDEEPQEEALPRGVVLR 211

RESULT 7
 A46462
 T cell activation antigen Tactile - human
 C:Species: Homo sapiens (man)
 C:Date: 18-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
 C:Accession: A46462
 R:Wang, P.L.; O'Farrell, S.; Clayberger, C.; Kravsky, A.M.
 J. Immunol. 148, 2600-2608, 1992
 A:Title: Identification and molecular cloning of Tactile. A novel human T cell acti-
 A:Reference number: A46462; MUID:92218864
 A:Accession: A46462
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-569 -WAN.
 A:Cross-references: GB:MBR282; NID:q18671; PIDN:AAA36562.1; PTD:q18672
 A:Note: sequence extracted from NCHI backbone (NCBIIN:94434, NCBIIP:94435)
 C:Keywords: Transmembrane protein

Query Match 6.14; Score 109; DB 2; Length 569;
 Best Local Similarity 20.74; Pred. No. 0.28;
 Matches 78; Conservative 71; Mismatches 145; Indels 85; Gaps 20;

QY 16 OLKGGCTIVNVYGVVKKFPKPPYLSKGTGYC-----SVVTIVD---OTNVKLTCLL----- 61
 Db 55 QMOWSKVTNKIDILAVYHPOV---GFCYACRPECESILVTETETENGSKMTLHLRNMSC 110
 QY 62 FSKRYEALPIYKNGDIVRFIRLKLQVYKHTQCTSSGFASLTFEGLTGAIPRTSS 120
 Db 111 SVSGRYECMLVLYPEIOTKIYNLLIOTHVTADEW-NSNHTLEIENOTLEIPCEQNSSS 169
 QY 121 KY-----ENFTTEDHKMVEALRVWASTIUMSPSWTLK-----LCDVQPMQYFD--- 163
 Db 170 KISSEFTYAWSVEDNGTQETL--ISQNLHLSNSTLLKDRVKRLGTDYRLHLSPVQIFDDGR 227

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 17:09:30 ; Search time 38.59 Seconds
(without alignments)
671.141 Million cell updates/sec

Title: US-09-816-248-15
Perfect score: 1776
Sequence: 1 MSLVPATNVIYTPNLQKGG.....LRPPCWSSVARSRLIAASTS 340

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	167.5	9.4	535	2 T38425	hypothetical prote
2	167	9.4	446	2 S35524	telomere-binding p
3	146.5	8.2	495	2 A41221	nucleonuclear alpha
4	142.5	8.0	460	2 S35525	telomere-binding p
5	135	7.6	493	2 S22581	telomere-binding p
6	130	7.3	251	2 T22839	hypothetical prote
7	109	6.1	569	2 A0462	T cell activation
8	109	6.1	587	2 T22006	hypothetical prote
9	103	5.8	1044	2 T00342	hypothetical prote
10	102.5	5.8	560	1 NCBPX6	exonuclease 46 (EC
11	102.5	5.8	4391	2 A38096	perlecan precursor
12	98	5.5	1198	2 D96723	hypothetical prote
13	95.5	5.4	474	2 T06149	hypothetical prote
14	94.5	5.3	462	2 B64443	hypothetical prote
15	93.5	5.3	400	1 JS0349	exonuclease (EC 4.
16	93.5	5.3	400	2 D85535	ATP-dependent dsbN
17	92.5	5.2	3707	2 S19252	heparan sulfate pr
18	90.5	5.1	610	2 G84306	biotin carboxylase
19	90.5	5.1	1030	2 F96763	hypothetical prote
20	90	5.1	361	2 E86692	N-acetyluracetyl-L
21	90	5.1	937	2 B86210	protein P2G5.6 [l
22	89.5	5.0	528	2 T32012	hypothetical prote
23	89.5	5.0	567	2 JN0459	transforming growt
24	89	5.0	421	1 F64205	thymidine phosphor
25	89	5.0	570	2 S07744	NADH dehydrogenase
26	89	5.0	683	2 B71325	conserved hypothet
27	88.5	5.0	1217	2 T21403	hypothetical prote
28	88	5.0	3788	2 T30851	lysosomal traffick
29	88	5.0	3796	2 T18514	lysosomal traffick

RESULT 2

30 87.5 4.9 553 2 T52362 hypothetical prote
31 87 4.9 4196 2 T41274 dynein heavy chain
32 86.5 4.9 318 2 A84466 hypothetical prote
33 86.5 4.9 592 2 S51371 transforming growt
34 86.5 4.9 1002 2 T46033 receptor protein k
35 86.5 4.9 1428 1 L5BYT2 DNA topoisomerase
36 86 4.8 899 2 C84765 hypothetical prote
37 85.5 4.8 574 2 A96612 hypothetical prote
38 85.5 4.8 861 2 T15903 protein kinase C h
39 85 4.8 426 1 N71249 hypothetical prote
40 85 4.8 2285 2 T12796 probable transglyc
41 84.5 4.8 334 2 F70384 hypothetical prote
42 84.5 4.8 507 2 S15476 transmembrane glye
43 84.5 4.8 749 2 H82691 topoisomerase IV s
44 84.5 4.8 905 1 RCBYS5 regulatory protein
45 84.5 4.8 1886 2 S04921 nuclear pore prote

ALIGNMENTS

RESULT 1
T38425
hypothetical protein SPAC26H5.06 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T38425
R:Oliver, K.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, September 1997
A:Reference number: Z21748
A:Accession: T38425
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-535 x0L>
A:Cross-references: EMBL:Z99126; P1DN:CA016192.1; GSPDB:GN00086; SPL:SPAC26H5.06
A:Experimental source: strain 972h-; cosmid c26H5
C:Genetics:
A:Gene: SPDB:SPAC26H5.06
A:Map position: 1
A:Introns: 44/1
C:Superfamily: Schizosaccharomyces hypothetical protein SPAC26H5.06

Query Match	9.4%	Score 167.5;	DB 2;	Length 535;
Best Local Similarity	24.4%	Pred. No. 2.7e-06;		
Matches 77;	Conservative 54;	Mismatches 118;	Indels 67;	Gaps 15;
QY 18	KGGTIVNVYGVVKKPKPPYLS-KGT-DYCSVVTIVDQT-----NVKLTCLLPSGNVEALP1	71		
DB 18	KKNTIVNLPGLVKRDETPSRQSLHGTCKDWVTTVYLMDPCTGTSSICGQIHLPKQCNLDLPV	77		
QY 72	LYKNGDIVERHRLKIQVKKETQGITSSGFA-----SLTEGTGAPLIPRTS-----	119		
DB 78	IKQVQPLALHQITKRSYKDRTOGLSKDQFRYALMDPDSNSKDKTLCQPMPLMKTGDK	137		
QY 120	-----SKYFNFTEDIKMVEALRWAS-----THMS-PSWTLKLCADVQPMQVFDLFC	166		
DB 138	EEQFALLINKINDEQTNKIKNGELLSTSSARONGTGLSPVSVFSLSLQITTHQRCSFYA	197		
QY 167	QLLGRKAEVDGASFLKLVMDTR-----TPPFS---WRVLQDLVLEGLSHIHLLQML	216		
DB 198	QVL-KTWSDKNFTLVYTDYTNELFPMSPYTSRWR-----CPFG-----RF	241		
QY 217	TIDLVDYDNIHVARS-LKVGSLRLYLSHITKLOSMNSSENQTMLSLEFLHKGTS--YGR	273		
DB 242	SRLCMLWDCHDPCYRNYIKEGDYVVKNVYRTKIDHLG-----YLECILHGDSSAKRYNM	294		
QY 274	GIRVLPSDSVDOLK 289			
DB 295	STEKVDSEPELNEIK 310			

Pr	27	-JUL-1994	9905-014555.1
Pr	28	-JUL-1994	9905-014556.1
Pr	02	-AUG-1994	9905-014638.6
Pr	02	-AUG-1994	9905-014638.8
Pr	02	-AUG-1994	9905-014638.9
Pr	03	-AUG-1994	9905-014700.3
Pr	04	-AUG-1994	9905-014720.4
Pr	04	-AUG-1994	9905-014730.2
Pr	05	-AUG-1994	9905-014730.2
Pr	05	-AUG-1994	9905-014726.0
Pr	06	-AUG-1994	9905-014730.3
Pr	06	-AUG-1994	9905-014741.6
Pr	09	-AUG-1994	9905-014742.9
Pr	09	-AUG-1994	9905-014734.5
Pr	10	-AUG-1994	9905-014817.1
Pr	11	-AUG-1994	9905-014831.9
Pr	12	-AUG-1994	9905-014834.1
Pr	13	-AUG-1994	9905-014836.5
Pr	13	-AUG-1994	9905-014846.8
Pr	16	-AUG-1994	9905-014936.8
Pr	17	-AUG-1994	9905-014917.5
Pr	18	-AUG-1994	9905-014944.2
Pr	20	-AUG-1994	9905-014972.2
Pr	20	-AUG-1994	9905-014972.3
Pr	20	-AUG-1994	9905-014996.2
Pr	20	-AUG-1994	9905-014996.2
Pr	23	-AUG-1994	9905-014993.0
Pr	25	-AUG-1994	9905-015003.6
Pr	26	-AUG-1994	9905-015008.4
Pr	27	-AUG-1994	9905-015106.5
Pr	27	-AUG-1994	9905-015106.6
Pr	27	-AUG-1994	9905-015108.0
Pr	30	-AUG-1994	9905-015130.3
Pr	31	-AUG-1994	9905-015143.8
Pr	01	-SEP-1994	9905-015133.0
Pr	07	-SEP-1994	9905-015236.3
Pr	10	-SEP-1994	9905-015307.0
Pr	13	-SEP-1994	9905-015375.8
Pr	15	-SEP-1994	9905-015401.8
Pr	16	-SEP-1994	9905-015403.9
Pr	20	-SEP-1994	9905-015477.9
Pr	22	-SEP-1994	9905-015513.9
Pr	23	-SEP-1994	9905-015565.9
Pr	24	-SEP-1994	9905-015645.6
Pr	26	-SEP-1994	9905-015645.6
Pr	29	-SEP-1994	9905-015679.6
Pr	04	-OCT-1994	9905-015711.7
Pr	05	-OCT-1994	9905-015775.3
Pr	06	-OCT-1994	9905-015766.5
Pr	07	-OCT-1994	9905-015802.9
Pr	08	-OCT-1994	9905-015832.3
Pr	12	-OCT-1994	9905-015836.9
Pr	13	-OCT-1994	9905-015929.3
Pr	13	-OCT-1994	9905-015929.4
Pr	13	-OCT-1994	9905-015949.9
Pr	14	-OCT-1994	9905-015932.9
Pr	14	-OCT-1994	9905-015933.0
Pr	14	-OCT-1994	9905-015933.1
Pr	14	-OCT-1994	9905-015936.3
Pr	14	-OCT-1994	9905-015936.8
Pr	18	-OCT-1994	9905-016074.4
Pr	21	-OCT-1994	9905-016074.7
Pr	21	-OCT-1994	9905-016076.6
Pr	21	-OCT-1994	9905-016087.0
Pr	21	-OCT-1994	9905-016087.4
Pr	21	-OCT-1994	9905-016088.0
Pr	22	-OCT-1994	9905-016098.1
Pr	22	-OCT-1994	9905-016098.7
Pr	23	-OCT-1994	9905-016140.4
Pr	25	-OCT-1994	9905-016140.6

Pk	2.6–6.07 ^a 1.949;	9.905–0.163 3.5;
Pp	2.6–6.027 ^a 1.949;	9.905–0.161 3.6;
Pr	2.6–6.07 ^a 1.949;	9.905–0.163 3.6;
Pb	2.6–6.07 ^a 1.949;	9.905–0.163 9.20;
Pk	2.6–6.07 ^a 1.949;	9.905–0.163 9.94;
Pb	2.6–6.07 ^a 1.949;	9.905–0.163 9.94;
Pb	2.9–6.027 ^a 1.949;	9.905–0.162 14.2;

Query Match 4.0%; Score 86; ID 21; Length 740;
 Best Local Similarity 21.2%; Pooled Ro. 5.1;
 Matches 72; Conservative 34; Mismatch 115; Indels 120; Gaps 15.

[illegible]

Search completed: April 9, 2002, 17:08:14
Job time: 592 sec

PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145921.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147494.
PR 09-AUG-1999; 99US-0147937.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148419.
PR 12-AUG-1999; 99US-0148441.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151063.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151448.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154036.
PR 20-SEP-1999; 99US-0154379.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 26-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157869.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158569.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.

ID AAW23594 standard; Protein: 3788 AA.
 AC AAW23594;
 DT 10-FEB-1998 (first entry)
 XX Murine Lyst1 long isoform.
 DE Lyst1; mouse; lysosomal trafficking regulator;
 KW Chediak-Higashi syndrome; CH syndrome; autoimmune disease; tumour;
 KW vaccine; therapy; diagnosis.
 XX Mus musculus.
 OS WO9728262-A1.
 PN 07-AUG-1997.
 PD 31-JAN-1997; 97WO-US01748.
 XX 23-DEC-1996; 96US-0034346.
 PR 01-FEB-1996; 96US-0011146.
 PR 20-DEC-1996; 96US-0033599.
 XX (UYFL) UNIV FLORIDA.
 PA Barbosa-Alleyne MDES, Kingsmore SF;
 PI WPI; 1997-402616/37.
 DR N-PSDB; AAT74199.
 XX Mammalian lysosomal trafficking regulators LYST1, LYST2 and
 PT LYST2 - useful to diagnose Chediak-Higashi syndrome
 XX Claim 3; Page 101-103; 237pp; English.
 XX This protein comprises the longer isoform of murine lysosomal
 CC trafficking regulator Lyst1, a protein that regulates desgranulation
 CC of lysosomes, late endosomes and acidic secretory granules
 CC primarily in leukocytes. Its sequence was deduced from a Lyst1
 CC cDNA clone (AAT74199) that had been identified from a beige (bg)
 CC mutation critical region yeast artificial chromosome. Alternative
 CC splicing also yields a shorter isoform (see AAW23595) of Lyst1.
 CC Homologous human sequences (see AAW23596 and AAW23597) have been
 CC identified that are associated with Chediak-Higashi syndrome (CHS).
 CC Murine Lyst1 and human LYST1 polypeptides (see AAW23598-98) can be
 CC expressed in recombinant host cells for use in raising antibodies,
 CC in vaccines and in the development of therapeutic approaches to
 CC treatment of autoimmune diseases and certain types of tumours. The
 CC existence of an animal model of CHS with a similar genetic lesion
 CC will assist efforts to develop novel therapies for this disease.
 XX Sequence 1788 AA;
 QWCV Match 5.0A; Score 88; DB 18; Length 3788;
 Best Local Similarity 18.8A; Pred. No. 43;
 Matches 81; Conservative 52; Mismatches 117; Indels 182; Gaps 17;
 QY 44 GSVVTVTDQTNVKTCLLFSGNVETALPIYKNGDILVHFHRLKIQVYKKFTQGITSSGFAS 107
 Db 1849 cnglsmlhqvkvkqk-----lqvfhllk----- 1872
 QY 104 LTFEGTIGAPIPTPTSSKYNFETEDHKWVEALR-----VWAS----- 141
 Db 1873 tllegccyeevvhvsehgefkldvshaiqdkllqellldwklnkacqgwetllaa 1932
 QY 142 -----TIMSPSWTLKLCQVPMQYEDLTCQLCK-----AE 173
 Db 1933 levllrvhqqqknikqlinahvwhlftcqlvqhregqltsmprevcrsfvkliae 1992
 QY 174 VDMA-----SFLKAVMDGTRT----- 189
 XX

Db 1993 vldspddlelltvlfnllavhptntvychmptnfvshldqklfekvqslaylrhs 2052
 QY 190 -----PPPSWRVLR-----QDLVLEGDLSHIHLQLNLTIDILVYDNHVIH 228
 Db 2053 ssggqafpspfivispsaftaappetsssnlypqlmaaqmvrslp-----afpctyp 2108
 QY 229 VARSCLKVGSFLRIYSLIHTKLQSM---NSENQTMLSLEFHLIGGTSTYGRGIRVLPESNSDV 285
 Db 2109 lraqkilaaslg-fsv-dkllqladanpekmllrpyalk--tskeafisscesaktv 2164
 QY 286 DOLKLDLESANLTAN-----QHSDFVCOSEPPDDSPNGVSL--RP---PCW 126
 Db 2165 cemea-llqahasagvscspfrprarvdhkvqteprtsdddsdpjdesyprtrpdmklgl 2223
 QY 327 SSVARSRLIAAS 338
 Db 2224 asfqsqstvas 2235
 RESULT 14
 AAG38741
 ID AAG38741 standard; Protein: 1002 AA.
 XX
 AC AAG38741;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein treatment. SEQ ID NO: 47839.
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX Arabidopsis thaliana.
 OS
 PN EPI033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 XX 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 18-MAY-1999; 99US-0134370.
 PR 19-MAY-1999; 99US-0134768.
 PR 20-MAY-1999; 99US-0134941.
 PR 21-MAY-1999; 99US-0135124.
 PR 24-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.

XX Forsyth KA, Ohlsen KL, Zyskind JW;
 XX WP1: 2001-457376/49.
 DR N-PSDB; AAH81418.
 XX Novel nucleic acids encoding proteins required for *Escherichia coli*
 PT proliferation, useful for screening for antimicrobial agents.
 PS Claim 19; Page 532-534; 596pp; English.
 XX The present invention describes a purified or isolated nucleic acid
 CC sequence (1) consisting essentially of one of the 93 nucleic acid sequences
 CC given in AAH81202 to AAH81294, where expression of the nucleic acid in a
 CC microorganism is capable of inhibiting proliferation of a microorganism.
 CC (1) have antibacterial and antibiotic activities, and can be used in
 CC gene therapy. Expression of (1) in a microorganism inhibits proliferation
 CC of the microorganism, and the manufactured antibiotic is useful for
 CC reducing the activity or level of a gene product required for
 CC proliferation of a microorganism in a subject, specifically humans. The
 CC nucleic acids that inhibit bacterial growth or proliferation can be used
 CC as antisense therapeutics for killing bacteria. In addition to
 CC therapeutic applications, the nucleic acid sequences complementary to
 CC sequences required for proliferation can be used as diagnostic tools.
 CC For example, nucleic acid probes complementary to proliferation-required
 CC sequences that are specific for particular species of microorganisms can
 CC be used as probes to identify particular species of microorganisms in
 CC clinical specimens. AAH81295 to AAH81487 encode the *Escherichia coli*
 CC proteins given in AAC98239 to AAC98431, and AAH81488 to AAH81491
 CC represent oligonucleotides, which are used in the exemplification of the
 CC present invention.
 XX Sequence 400 AA:
 SO
 Query Match 5.3%; Score 93.5; DB 22; Length 400;
 Best Local Similarity 21.2%; Pred. No. 0.25;
 Matches 69; Conservative 38; Mismatches 119; Indels 99; Gaps 14;
 QY 34 PNYLSKCTDYGVVTVDTQNVKLTCLFSQNEVALP1YKNCJIVRPHKIKOVYKKRP 93
 DB 55 PNYAR-LIYRIVLVNLIQGLHVL--QGNDSVALINCSIDJMA-----IHL 102
 QY 94 QGTTSSGFASLTP---KGTLCAL11P-----KTSRYFNTTEDIHKEVLEAKWAS 141
 DB 103 TVAASQHPQPIPRDQPPQAVICP1PIPRDQIPLSQGLNICKQGLDADIDYQ 162
 QY 142 THMSPSMTLKLCDVOPMOYFDLTGCLIGRAVYDAS---FLIKVMDGTPPTSMFVL 198
 DB 163 QHYADE---CKLIGQDQPLPIJ-----ALQHLLVQASKSQAVIDIYQLDAPQPIFP 214
 QY 199 ODVLVLEGDSIHTHRLONLTIDILVYDNHIVAVASLAKVCSILKILTKLQSNSENCIM 258
 DB 215 ADYIA---Iqhlhraq-----Ilgmehvryqcs-----P 241
 QY 259 LSEPLHIGTSGRGRCIVLPESNSPDV-----LKKDLPSANI/TANOHSD 304
 DB 242 IPISTCECKSKSY---VHVLVLSNCKIESVENIIVNVLGPMUAVLKQDASILQALCEQWID 298
 QY 305 VTCSEPTDPSFVNCVASLRPCWSSV 329
 DB 299 -----VSQPPVWIDJ 309
 RESULT 12
 AAH47250
 ID AAH47250 standard; Protein; 653 AA.
 XX
 AC AAH47250;
 XX
 DT 18-JUL-2001 (first entry)
 XX
 DE Human PP7.

XX Mitochondrial calcium-binding protein; p95.6/YNF52; calcium;
 KW drug screening; sequence extraction; PT7; calcium binding motif.
 XX
 OS Homo sapiens.
 XX MO200130H40-A2.
 PN
 XX 03-MAY-2001.
 PD
 XX 26-OCT-2000; 2000MO-US29787.
 PT
 XX 26-OCT-1999; 99MS-0427867.
 PR 26-OCT-1999; 99MS-0427895.
 PR 07-SEP-2000; 2000HS-0657253.
 XX (MITO-) MITOKOR.
 PA
 XX Murphy AN, Fahy ED, Wiley SR, Cleverger W;
 PT WP1: 2001-108616/42.
 DR N-PSDB; AAC85804.
 XX New Ca²⁺-dependent cytochrome mitochondrial calcium-binding protein,
 PT p95.6/YNF52, identified by protein motif database searching, useful for
 PT localizing intracellular calcium ions and in drug screening assays.
 XX Example 2; Fig 1; 176pp; English.
 XX This sequence shows a protein containing calmodulin/calcium
 CC binding motifs which was identified using the method of the invention.
 CC The method comprises a computer system for extracting information
 CC from biological sequence databases, which receives a motif that defines
 CC a desired polypeptide sequence. The computer searches the database
 CC based on a query formulated from the motif to obtain a set of
 CC polypeptide sequences which are then filtered and a report of the
 CC filtered set of polypeptide sequence information is then displayed.
 CC The computer system is useful for identifying recombinant cytochromes
 CC p95.6/YNF52 and other mitochondrial proteins involved in binding,
 CC transport or other regulation of intracellular calcium which provides
 CC targets for screening assays that identify agents capable of affecting
 CC a mitochondrial activity.
 XX Sequence 653 AA:
 SO
 Query Match 5.0%; Score 88.5; DB 22; Length 653;
 Best Local Similarity 21.5%; Pred. No. 2.2;
 Matches 63; Conservative 32; Mismatches 93; Indels 105; Gaps 15;
 QY 120 SKYFNFTTEP---HKNVEALR---WVASTHMSPSMTLKLCDVQ-----PMQYFDLTC 166
 DB 60 STLSIMENYLIHKEGLINQSLTSCQDMRDWDYDSDVDSYNGRP1PIPLECT 119
 QY 167 -----QLCKRAVEDGASFLIKVMDGTPPTF-----PSMRVLI-----OD 200
 DB 120 DIDILIEAFKEQGLIHABHYLEVJTEKKVILKQMPFHLIQSPSKSEVLEIGDILHKGKID 179
 QY 201 L-----VLKCDL-----SHIHLONLTIDILVYDNHIVAR----- 231
 DB 180 LILLYKNQIPSEIRNYPVINDIIVTRKNSIELIMLEVSILYPPNDIHLINQNHEDIMM 239
 QY 232 SLKVC---STLKLISLALTK---LQSMNS-----ENQYMLSLSEFHLNGCTSYGKGI 275
 DB 240 NITRYATKEELIKKYLHQRILIQILEVYAWPIQLIVQNEILV-----IlggIscetdI 294
 QY 276 RVL-----PESNSIVPOLKRDLESANLT/ANOHSDVLCOSEPDD 313
 DB 295 NIIHYVERIKMKSVLIPPLECTINIDITLDSKHNKYGVTINAHRIKLTNSPCL 346
 RESULT 13
 AAW23594

PK 26-OCT-1999; 9905-0161361.
PK 28-OCT-1999; 9905-0161920.
PK 28-OCT-1999; 9905-0161992.
PK 28-OCT-1999; 9905-0161993.
PK 29-OCT-1999; 9905-0162142.

Query Match 5.5%; Score 98; Bk 21; Length 1058;
Best Local Similarity 21.4%; Pred. No. 0.45;
Matches 65; Conservative 36; Mismatches 119; Indels 84; Gaps 15;

QY 50 VDDGNVKKLQCLLP---SCNVEALLIYKNCNLIYVPHLKLQVKKKIKQCIYSSGCFASITP 106
DB 629 meeqnfpdtejckdsqeqvqspavyt-qhdtsiqesklqkums----lde-stahnl 682
QY 107 ECTTGA-P11PTSSSKYNNFTTEHDKMVEALVKVMASTHMSPSW--TILKICLVQHPQVY 162
DB 683 cdeclaudcpkqvkvkpsyyvalgsdhsywlclckwrltyvrtqfplalalqalqpr--- 739
QY 163 DLTCGLLKKEVNGCASFLKLVMDGCTKTPPSKIVLQDVLAKDLSLHHKGLQNLTDILV 222
DB 740 clmknjldkadm-----qvw-----dklqesclstslwtlq----- 769
QY 223 YDNHIVAKSLKGVFLKRYSLHTR--LOSMSNPOPMLSLEPHLDCCTSYGRCHEVLPES 281
DB 770 -----slelws-----wslhvpvlsstlsdcv-----lmsscyydhnaqkykcy 809
QY 282 NSNPVQLKRD-----LESANLTPANQHSNVLCSEHVDNSPNCV-SLKPQWS 327
DB 810 rsklsrlkkkmsrscqfryckkljclanllvcemkrlkqrst---rlforlqlypsvsk 866
QY 328 SVAK 331
DB 867 dfmr 870

RESULT 16

AAG30850
ID AAG30850 standard; Protein; 1198 AA.

AC AAG30850;

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SRO 1P No: 66955.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

XX EPI03405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 9905-0121825.

PK 05-MAR-1999; 9905-0123180.

PK 09-MAR-1999; 9905-0123548.

PK 23-MAR-1999; 9905-0125788.

PK 25-MAR-1999; 9905-0126264.

PK 29-MAR-1999; 9905-0126785.

PK 01-APR-1999; 9905-0127462.

PK 06-APR-1999; 9905-0128234.

PK 08-APR-1999; 9905-0128714.

PK 16-APR-1999; 9905-0129845.

PK 19-APR-1999; 9905-0130077.

PK 21-APR-1999; 9905-0130449.

PK 23-APR-1999; 9905-0130510.

PK 28-APR-1999; 9905-0130891.

PK 30-APR-1999; 9905-0131449.

PK 30-APR-1999; 9905-0132048.

PK 30-APR-1999; 9905-0132407.
PK 04-MAY-1999; 9905-0132464.
PK 05-MAY-1999; 9905-0132485.
PK 06-MAY-1999; 9905-0132486.
PK 06-MAY-1999; 9905-0132487.
PK 07-MAY-1999; 9905-0132863.
PK 11-MAY-1999; 9905-0134256.
PK 14-MAY-1999; 9905-0134218.
PK 14-MAY-1999; 9905-0134219.
PK 14-MAY-1999; 9905-0134221.
PK 14-MAY-1999; 9905-0134370.
PK 18-MAY-1999; 9905-0134764.
PK 18-MAY-1999; 9905-0134941.
PK 20-MAY-1999; 9905-0135124.
PK 21-MAY-1999; 9905-0135353.
PK 24-MAY-1999; 9905-0135629.
PK 25-MAY-1999; 9905-0136021.
PK 27-MAY-1999; 9905-0136392.
PK 28-MAY-1999; 9905-0136782.
PK 01-JUN-1999; 9905-0137222.
PK 03-JUN-1999; 9905-0137528.
PK 04-JUN-1999; 9905-0137502.
PK 07-JUN-1999; 9905-0137724.
PK 08-JUN-1999; 9905-0138094.
PK 10-JUN-1999; 9905-0138540.
PK 10-JUN-1999; 9905-0138847.
PK 14-JUN-1999; 9905-0139119.
PK 16-JUN-1999; 9905-0139452.
PK 17-JUN-1999; 9905-0139453.
PK 18-JUN-1999; 9905-0139492.
PK 18-JUN-1999; 9905-0139455.
PK 18-JUN-1999; 9905-0139456.
PK 18-JUN-1999; 9905-0139457.
PK 18-JUN-1999; 9905-0139458.
PK 18-JUN-1999; 9905-0139459.
PK 18-JUN-1999; 9905-0139460.
PK 18-JUN-1999; 9905-0139461.
PK 18-JUN-1999; 9905-0139462.
PK 18-JUN-1999; 9905-0139463.
PK 18-JUN-1999; 9905-0139750.
PK 18-JUN-1999; 9905-0139763.
PK 21-JUN-1999; 9905-0139817.
PK 22-JUN-1999; 9905-0139899.
PK 23-JUN-1999; 9905-0140353.
PK 23-JUN-1999; 9905-0140354.
PK 24-JUN-1999; 9905-0140695.
PK 28-JUN-1999; 9905-0140823.
PK 29-JUN-1999; 9905-0140991.
PK 30-JUN-1999; 9905-0141287.
PK 01-JUL-1999; 9905-0141842.
PK 01-JUL-1999; 9905-0142154.
PK 02-JUL-1999; 9905-0142055.
PK 06-JUL-1999; 9905-0142390.
PK 08-JUL-1999; 9905-0142803.
PK 09-JUL-1999; 9905-0142920.
PK 12-JUL-1999; 9905-0142977.
PK 13-JUL-1999; 9905-0143542.
PK 14-JUL-1999; 9905-0143624.
PK 15-JUL-1999; 9905-0144005.
PK 16-JUL-1999; 9905-0144085.
PK 16-JUL-1999; 9905-0144086.
PK 19-JUL-1999; 9905-0144325.
PK 19-JUL-1999; 9905-0144331.
PK 19-JUL-1999; 9905-0144332.
PK 19-JUL-1999; 9905-0144333.
PK 19-JUL-1999; 9905-0144334.
PK 19-JUL-1999; 9905-0144335.
PK 20-JUL-1999; 9905-0144352.
PK 20-JUL-1999; 9905-0144632.
PK 20-JUL-1999; 9905-0144864.
PK 21-JUL-1999; 9905-0144814.
PK 21-JUL-1999; 9905-0145086.

PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140351.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159291.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159129.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	26-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.

PK	11-AUG-1999;	990S-0148319.
PK	12-AUG-1999;	990S-0148341.
PK	13-AUG-1999;	990S-0148565.
PK	13-AUG-1999;	990S-0148684.
PK	16-AUG-1999;	990S-0149368.
PK	17-AUG-1999;	990S-0149175.
PK	18-AUG-1999;	990S-0149426.
PK	20-AUG-1999;	990S-0149722.
PK	20-AUG-1999;	990S-0149723.
PK	20-AUG-1999;	990S-0149929.
PK	23-AUG-1999;	990S-0149902.
PK	23-AUG-1999;	990S-0149930.
PK	25-AUG-1999;	990S-0150566.
PK	26-AUG-1999;	990S-0150884.
PK	27-AUG-1999;	990S-0151065.
PK	27-AUG-1999;	990S-0151066.
PK	27-AUG-1999;	990S-0151080.
PK	30-AUG-1999;	990S-0151303.
PK	31-AUG-1999;	990S-0151438.
PK	01-SEP-1999;	990S-0151930.
PK	07-SEP-1999;	990S-0152363.
PK	10-SEP-1999;	990S-0153070.
PK	13-SEP-1999;	990S-0153758.
PK	15-SEP-1999;	990S-0154018.
PK	16-SEP-1999;	990S-0154039.
PK	20-SEP-1999;	990S-0154779.
PK	22-SEP-1999;	990S-0155139.
PK	23-SEP-1999;	990S-0155486.
PK	24-SEP-1999;	990S-0155659.
PK	28-SEP-1999;	990S-0156458.
PK	29-SEP-1999;	990S-0156596.
PK	04-OCT-1999;	990S-0157117.
PK	05-OCT-1999;	990S-0157753.
PK	06-OCT-1999;	990S-0157865.
PK	07-OCT-1999;	990S-0158029.
PK	08-OCT-1999;	990S-0158232.
PK	12-OCT-1999;	990S-0158365.
PK	13-OCT-1999;	990S-0159293.
PK	13-OCT-1999;	990S-0159294.
PK	13-OCT-1999;	990S-0159295.
PK	14-OCT-1999;	990S-0159329.
PK	14-OCT-1999;	990S-0159330.
PK	14-OCT-1999;	990S-0159331.
PK	14-OCT-1999;	990S-0159637.
PK	14-OCT-1999;	990S-0159638.
PK	18-OCT-1999;	990S-0159584.
PK	21-OCT-1999;	990S-0160741.
PK	21-OCT-1999;	990S-0160767.
PK	21-OCT-1999;	990S-0160768.
PK	21-OCT-1999;	990S-0160770.
PK	21-OCT-1999;	990S-0160814.
PK	21-OCT-1999;	990S-0160815.
PK	22-OCT-1999;	990S-0160980.
PK	22-OCT-1999;	990S-0160981.
PK	22-OCT-1999;	990S-0160989.
PK	25-OCT-1999;	990S-0161404.
PK	25-OCT-1999;	990S-0161405.
PK	25-OCT-1999;	990S-0161406.
PK	26-OCT-1999;	990S-0161359.
PK	26-OCT-1999;	990S-0161360.
PK	26-OCT-1999;	990S-0161361.
PK	28-OCT-1999;	990S-0161920.
PK	28-OCT-1999;	990S-0161992.
PK	28-OCT-1999;	990S-0161993.
PK	29-OCT-1999;	990S-0162142.

Query Match 5.5%; Score 98; DO 21; Length 955;
Best Local Similarity 21.4%; Pred. No. 0.48;
Matches 65; Conservative 36; Mismatches 119; Indels 84; Gaps 15;

50 VDOTNVLTLTLF--SCNTEALPIYKNCIDYRPHIKIOVKKKKCGCTTSSGVSLSLP 106
526 meenlpfclclckdsegqsgsvpavyj-qhldstqgsklgkums-----ldstalnll 579

QY	107	ECGTLGA--P11PTSSKRYNFETTHDKRMVLAJVVMASTHMSYSW--TALKCTVOPMAYVF 162
DB	580	cdad-laud-pakpvcpkpvcgrygvdtsdhwxy1110kwtctvy1010pualtdadp- 640
QY	163	DLTCGLACNAEYVCASPLAKVMIWCTRTHTTSSKVALYQDMLAKDLSHDKQNTLTDLLV 222
DB	637	clnkulldkdam-----qw-----dklarsclastwt 606
QY	223	YDNIHVAVKSLKVGCSPLRYSLATRK--LQSMNSKNGPMLSLFPHJGCTSYGCTHVLJPS 281
DB	667	-----slctws-----wslhnpvls1sdctv- ---lhmsvgydaupkykq 706
QY	282	NSHVDGLKRP-----LQZANLTVANQSHVYFCQETHPDQSPNGV SLAFPSWGS 327
DB	707	ckklsllkkkmsa sgcltgykklldgnlltvcemkllkaur- ---llhlsqspgawk 764
QY	328	SVAR 341
DB	764	dlm 767
RESULT 9		
DB	AAC30851	
XX	AAC30851	standard; proteing 1074 aa.
AC	AAC30851	
XX		
DT	17-OCT-2000	(first entry)
XX		
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 36950.	
XX		
KW	Protein identification: signal transduction pathway; metabolic pathway;	
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	
KW	terminator sequence.	
OS	Arabidopsis thaliana.	
XX		
PN	EP1033405-A2.	
XX		
PD	06-SEP-2000.	
XX		
PF	25-FEB-2000; 2000EP-0301439.	
XX		
PK	25-FEB-1999;	990S-0121825.
PK	05-MAR-1999;	990S-0123180.
PK	09-MAR-1999;	990S-0123548.
PK	23-MAR-1999;	990S-0125788.
PK	25-MAR-1999;	990S-0126264.
PK	29-MAR-1999;	990S-0126785.
PK	01-APR-1999;	990S-0127462.
PK	06-APR-1999;	990S-0128234.
PK	08-APR-1999;	990S-0128714.
PK	16-APR-1999;	990S-0129845.
PK	19-APR-1999;	990S-0130077.
PK	21-APR-1999;	990S-0130449.
PK	23-APR-1999;	990S-0130510.
PK	23-APR-1999;	990S-0130819.
PK	28-APR-1999;	990S-0131449.
PK	30-APR-1999;	990S-0132048.
PK	30-APR-1999;	990S-0132407.
PK	04-MAY-1999;	990S-0132484.
PK	05-MAY-1999;	990S-0132485.
PK	06-MAY-1999;	990S-0132486.
PK	06-MAY-1999;	990S-0132487.
PK	07-MAY-1999;	990S-0132863.
PK	11-MAY-1999;	990S-0144256.
PK	14-MAY-1999;	990S-0144218.
PK	14-MAY-1999;	990S-0134219.
PK	14-MAY-1999;	990S-0134221.
PK	14-MAY-1999;	990S-0134370.
PK	18-MAY-1999;	990S-0134760.
PK	19-MAY-1999;	990S-0134941.

DE Usurin-gamma polypeptide.
 XX
 XX Usurin-alpha: death effector domain; DE2; prodomain; usurin-beta;
 KW usurin-gamma; procaspase-8; CD95; apoptosis; cancer; immunosuppressive;
 KW caspase; cytostatic; antiParkinsonian; antidiabetic.
 XX
 OS Homo sapiens.
 XX
 XX W0200003023-A1.
 XX
 XX 20-JAN-2000.
 XX
 XX 07-JUL-1999; 99KO-CA00615.
 XX
 XX 08-JUL-1998; 98US-0092005.
 XX
 XX (MERK) MERCK PROSST CANADA INC.
 XX
 XX Nicholson DM, Kasper DM, Xanthopoulos S, Roy S;
 DE WPI: 2000-160929/14.
 DE N-PSDH; AAZ56989.
 XX
 XX Novel recombinant DNA molecules and polypeptides for treating apoptosis
 PT mediated diseases e.g. autoimmune diseases, cancer and Parkinson's
 PT disease.
 XX
 XX Claim 6; Fig 11h; 69pp; English.
 XX
 XX The invention provides recombinant nucleic acid molecules encoding
 CC usurin-alpha (lacking the first death effector domain (DE1) or its
 CC prodomain), usurin-beta or usurin-gamma. Usurin polypeptides are
 CC useful for in vitro and in vivo identification of usurin-procaspase-8
 CC interaction inhibitor. Usurin is useful as modulator of the sensitivity
 CC of cells to CD95(Fas/Apo-1) mediated apoptosis. Modulation of apoptosis
 CC is useful for treating diseases like autoimmune diabetes, cancer and
 CC Parkinson's disease. Activators and inhibitors of usurin-procaspase-8
 CC interaction are also useful for treating various diseases mediated by
 CC apoptosis. Usurin provides an attractive model for modulating caspase
 CC activation. Sensitivity of cells bearing CD95(Fas/Apo-1) receptor can
 CC be regulated at several levels in the presence of usurin, conferring
 CC resistance to Fas-ligand cell death. The present sequence represents
 CC the usurin-gamma polypeptide.
 CC
 XX
 XX Sequence 291 AA:
 SO
 Query Match 5.8%; Score 103; DB 21; Length 291;
 Best Local Similarity 25.0%; Pred. No. 0.016;
 Matches 49; Conservative 30; Mismatches 73; Indels 44; Gaps 7;
 QY 169 LGKAEVNGASFLKLVNDG-----TKTPPSKWLJLDLY-----LEKDLSHIRLON 215
 DB 104 LKSDVSLILFKDYMGKQKSKSKSILDLWELKINLVAPQDLLEKLNIRI-D 162
 QY 216 LTTDLIVYDNIHIVAVSLKVGSLFKIYSLITKLSQMSNSQMTLSLEPIHIGCTVCKGI 275
 DB 163 LKLLKQKYSVVGAGT-----SYTNVIGAILIKSKLKPSSNLTILHMFSSQPI 212
 QY 276 RVLPSNSDVLQKDLKDLNANLTANQHSNVLCOSEHDDSP-----PNCVSL-----RNP 324
 DB 213 K-----eqLqagqepvkkiskgscaL lpsqipwcyrykmskspqlqclldelqnet 263
 QY 325 GNSVAKSRKLIAASTS 340
 DB 264 qvsamagsgltaists 279
 RESULT 7
 AAB31889
 ID AAB31889 standard. Protein: 4393 AA.
 AC AAB31889;

XX
 XX 15-MAY-2001 (first entry)
 DE
 XX Amino acid sequence of a human protein.
 DE
 XX
 KW human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW hematoid polynthritis; lupus erythematosus; gene therapy.
 XX
 OS Homo sapiens.
 XX
 XX W0200105422-A2.
 XX
 XX 25-JAN-2001.
 XX
 XX 17-JUL-2000; 2000MO-PRO2057.
 XX
 XX 15-JUL-1999; 99PR-0009472.
 XX
 XX (INMR) BIOMERIEUX STELLAYS.
 XX
 XX Kocxkin D, Kollie H, Charles M, Marcus C, Santoro L, Perron H;
 DE WPI: 2001-159475/16.
 DE
 XX Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specific
 PT polypeptides or related nucleic acid or ligand.
 XX
 XX Claim 1; Page 136-152; 20pp; French.
 XX
 XX The present sequence represents a human protein, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms
 CC and phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritits and lupus erythematosus, including use as vaccines and
 CC in gene therapy (expression of sense or antisense sequences). They can
 CC also be used to assess efficacy of potential therapeutic agents,
 CC particularly compounds that reduce or inhibit toxicity towards glial
 CC cells.
 CC
 XX
 XX Sequence 4393 AA:
 SO
 Query Match 5.8%; Score 102.5; DB 22; Length 4393;
 Best Local Similarity 24.1%; Pred. No. 1.5;
 Matches 84; Conservative 39; Mismatches 140; Indels 85; Gaps 14;
 QY 48 TIVDOTNRKLTCLFSCNTEALALIVKNCJL-----VRPIHKLI-----QVYRKIE 92
 DB 2257 LVaeqql ldlsevv-aqgaqavt wykrgas lparhayrasfYl lqspadaqgyctr 2415
 QY 93 TCGTSSSCASALTFPGCTLGA-----P1-IPRTSSKVFNNFTTED----- 129
 DB 2316 suqmeas--llevtdlqdanluybaqst qptleppssqvaegqlldncvpagshag 2373
 QY 130 --HKMEVLAIRKMASTHHSPTSLKLCIDVQHPQYFDLTCLLKAKEVDCASFLKVMK 186
 DB 2374 vclwhkrqslpvrhqlbus-----lllyqaspadasqeyvcrvJassvpldeasvvllepa 2429
 QY 167 TRTP-----PSSKWLJLDLYLKKLSHILKLNLTDLVLYDN-HVNV-----AK 231
 DB 2430 gsvpelavlrvvtl-----cvsstsqvaagqtlldlnc lvaqgaqayt wlkpaas lpar 2487


```

Db 181 lkwtdgtrtpfswrvliqdlvlegdlsghrlqnlctidilvydnhvnhvarsikvgsflr 240
QY 241 lYSLHTKLSQMSNSENQMLSLFHLHGTSYGRGIRVLPSNSDSDVOLLKDLKLESANLTAN 300
Db 241 lYSLHTKLSQMSNSENQMLSLFHLHGTSYGRGIRVLPSNSDSDVOLLKDLKLESANLTAN 300
QY 301 QHSDVICQSEPDSPNGVSL 321
Db 301 qhsdvicqsepdspngvsgv 321

RESULT 3
AAB93478
ID AAB93478 standard; Protein: 634 AA.
XX AC AAB93478;
XX DT 26-JUN-2001 (first entry)
XX Human protein sequence SEQ ID NO:12761.
XX KW Human; primer: detection; diagnosis; antisense therapy; gene therapy.
XX OS Homo sapiens.
XX PN EP1074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isozaki T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Saito Y, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX Claim 8; SEQ ID 12761; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dt primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB93893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification

```

```

CC of the present invention.
XX SQ Sequence 634 AA;
XX Query Match 93.5%; Score 1660; DB 22; Length 634;
XX Best Local Similarity 98.8%; Prod. No. 2.9a-171;
XX Matches 317; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 MSLVPATNYLYTPLNOLKGDTLVNVYGVKFFKPPYLSKGTDCYCVVTIVDQTNVKTCL 60
Db 1 mslvpatnylytplnqlkgdtlvnvvgvkvklfkppylskgtcyssvtlvdqnlvklcl 60
QY 61 LFSGNYEALPIIYKNGDIVERFHLKTOVYKKEGTTSSGFSALTFEGTLCADIPRTSS 120
Db 61 lfsqnyealpitykngdivrfhrlkiqvykktggtssgfasltfegtlqapilptss 120
QY 121 KYENFTTEDHKKVEALRVWASTHMSPSWTLKLCDDVOPMOYFDLTCCOLLGKAEVDGASFL 180
Db 121 kyntfttedhkhmvealrvwasthmspswtllk lcdvqpmqyldlccqlgkaevdgastl 180
QY 181 LKWDGTRTPFPFSRVLITDVLVLEGLDLSHHRHQNLTIDILVYDNHVHVARSLKVGSLR 240
Db 181 lkwdgtrtpfswrvliqdlvlegdlsghrlqnlctidilvydnhvnhvarsikvgsflr 240
QY 241 lYSLHTKLSQMSNSENQMLSLFHLHGTSYGRGIRVLPSNSDSDVOLLKDLKLESANLTAN 300
Db 241 lYSLHTKLSQMSNSENQMLSLFHLHGTSYGRGIRVLPSNSDSDVOLLKDLKLESANLTAN 300
QY 301 QHSDVICQSEPDSPNGVSL 321
Db 301 qhsdvicqsepdspngvsgv 321

RESULT 4
AAB95120
ID AAB95120 standard; Protein: 634 AA.
XX AC AAB95120;
XX DT 26-JUN-2001 (first entry)
XX Human protein sequence SEQ ID NO:17113.
XX KW Human; primer: detection; diagnosis; antisense therapy; gene therapy.
XX OS Homo sapiens.
XX PN EP1074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isozaki T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Saito Y, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX Claim 8; SEQ ID 17113; 2537pp + CD ROM; English.

```

(HSE-) HSE INC.

Tang Y.T., Liu C., Asundi V., Chen R., Ma Y., Qian X., Ren F., Wang D.;
 Wang J., Wang Z., Wehrman T., Xu C., Xu A.J., Yang Y., Zhang J.;
 Zhao Q., Zhou P., Goodrich K., Dermanic RT;
 WI: 2001-442253/47.
 N-PSDH: AA159281.

Novel nucleic acids and polypeptides, useful for treating disorders
 such as central nervous system injuries -

Example 5: SEQ ID NO 3270; 10078pp; English.

The invention relates to human nucleic acids (AA157748-AA161369) and
 the encoded polypeptides (AA38642-AA42213) with neurotrophic,
 immunosuppressant and cytostatic activity. The polypeptides or polynucleotide
 in gene therapy. A composition containing a polypeptide of the peripheral nervous
 system, the invention may be used to treat diseases of the peripheral nervous
 system, such as peripheral nervous injuries, peripheral neuropathy and
 localized neuropathies and central nervous system diseases, such as
 Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 utilization of the activities such as: immune system suppression, the
 activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
 and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 assays for receptor activity, arthritis and inflammation, leukemias and
 C.N.S. disorders.

Note: The sequence data for this patent did not form part of the printed
 specification.

Sequence 634 AA:

Query Match 93.5%; Score 1660; Dh 22; Length 634;
 Best Local Similarity 98.0%; Pred. No. 2,96-171;
 Matches 317; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

1 MSLVATNTVITPLNOLKGTIVNVGVVAFKPPYLSKGTGYCVVTVDOTNKLCTL 60
 |||||||
 1 MSIVPALNYIYLPINQIKQGLVNVYGVVKKIPYISKGLGYCVVIVDGLNKLCT 60
 |||||||
 61 LPSGVNVALPIYKNGDIVFHRILKIOVYKKEVGGITSSGFASLPEEGTAAPIPTSS 120
 |||||||
 61 LSGNYEALPIYKNGDIVFHRILKIOVYKKEVGGITSSGFASLPEEGTAAPIPTSS 120
 |||||||
 121 KYFNTTEDIKMWVALLRWASTHMSFSTLLKLCVOPMOYFDLTCOLLGKAEVDCASFL 180
 |||||||
 121 KYINFTEDEHKMVEALRWASTHMSFSTLLKLCVOPMOYFDLTCOLLGKAEVDCASFL 180
 |||||||
 181 LKVMGCTPTPTPSKRVLIQDLVLEGLSHIHRLONLTDILVYDNHIVAKSLKVSFLR 240
 |||||||
 181 LKVGQGLTPTPTPSKRVLIQDLVLEGLSHIHRLONLTDILVYDNHIVAKSLKVSFLR 240
 |||||||
 241 IYSLHTKLQSNMSENQIMLSLEFHLHGCTSYGKCIKVLIPESNSVDOLKEDLESANLTAN 300
 |||||||
 241 IYSLHTKLQSNMSENQIMLSLEFHLHGCTSYGKCIKVLIPESNSVDOLKEDLESANLTAN 300
 |||||||
 301 QHSNVTICQSEFDSPKCVSL 321
 |||||||
 301 QHSNVTICQSEFDSPKCVSL 321
 |||||||

RESULT 2
 AAB92742
 ID AAB92742 standard; Protein: 634 AA.
 AC AAB92742;
 XX 26-JUN-2001 (first entry)
 XX Human Protein sequence SEQ ID NO:11108.

Human; primer: detection; diagnosis; antisense therapy; gene therapy.

Homo sapiens.
 EF1074617-A2.
 07-FEB-2001.
 28-JUL-2000; 2000EF-0116126.
 29-JUL-1999; 95JP-0248036.
 27-AUG-1999; 99JP-0300253.
 11-JAN-2000; 2000JP-0118776.
 02-MAY-2000; 2000JP-0183767.
 09-JUN-2000; 2000JP-0241809.
 (HELIX-) HELIX RES. INST.

Otsu T., Itochi T., Nishikawa T., Hayashi K., Saito K., Yamamoto J.;
 Ishii S., Sugiyama T., Wakamatsu A., Nagai K., Otsuki T;
 WI: 2001-338749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602
 full-length cDNAs defined in the specification, and for the detection
 and/or diagnosis of the abnormality of the proteins encoded by the
 full-length cDNAs -

Claim B; SEQ ID 11108; 2547pp; CD ROM; English.

The present invention detects primer sets for synthesizing 5602
 full-length cDNAs defined in the specification, where a primer set
 comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 to the complementary strand of a polynucleotide which comprises one of
 the 5602 nucleotide sequences defined in the specification, where the
 oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 of an oligonucleotide comprising a sequence complementary to the
 complementary strand of a polynucleotide which comprises a 5'-end
 sequence and an oligonucleotide comprising a sequence complementary to a
 polynucleotide which comprises a 3'-end sequence, where the
 oligonucleotide comprises at least 15 nucleotides and the combination of
 the 5'-end sequence/3'-end sequence is selected from those defined in
 the specification. The primer sets can be used in antisense therapy and
 in gene therapy. The primers are useful for synthesizing polynucleotides,
 particularly full-length cDNAs. The primers are also useful for the
 detection and/or diagnosis of the abnormality of the proteins encoded by
 the full-length cDNAs. The primers allow obtaining of the full-length
 cDNAs easily without any specialized methods. AAH03166 to AAH13628 and
 AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 represent oligonucleotides, all of which are used in the exemplification
 of the present invention.

Sequence 634 AA:

Query Match 93.5%; Score 1660; Dh 22; Length 634;
 Best Local Similarity 98.0%; Pred. No. 2,96-171;
 Matches 317; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

1 MSLVATNTVITPLNOLKGTIVNVGVVAFKPPYLSKGTGYCVVTVDOTNKLCTL 60
 |||||||
 1 MSIVPALNYIYLPINQIKQGLVNVYGVVKKIPYISKGLGYCVVIVDGLNKLCT 60
 |||||||
 61 LPSGVNVALPIYKNGDIVFHRILKIOVYKKEVGGITSSGFASLPEEGTAAPIPTSS 120
 |||||||
 61 LSGNYEALPIYKNGDIVFHRILKIOVYKKEVGGITSSGFASLPEEGTAAPIPTSS 120
 |||||||
 121 KYFNTTEDIKMWVALLRWASTHMSFSTLLKLCVOPMOYFDLTCOLLGKAEVDCASFL 180
 |||||||
 121 KYINFTEDEHKMVEALRWASTHMSFSTLLKLCVOPMOYFDLTCOLLGKAEVDCASFL 180
 |||||||
 181 LKVMGCTPTPTPSKRVLIQDLVLEGLSHIHRLONLTDILVYDNHIVAKSLKVSFLR 240
 |||||||
 181 LKVGQGLTPTPTPSKRVLIQDLVLEGLSHIHRLONLTDILVYDNHIVAKSLKVSFLR 240
 |||||||

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 17:08:10 ; Search time 58.75 seconds
(without alignments)
428.679 Million cell updates/sec

Title: US-09-816-248-15
Perfect score: 1776
Sequence: 1 MSLLPATNVIYPLNQLKLG.....LRPPGWSVAKSRLLIAASTS 340

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SIDSB/qcdata/qeneseq/qeneseq/AA1980.DAT.*
2: /SIDSB/qcdata/qeneseq/qeneseq/AA1981.DAT.*
3: /SIDSB/qcdata/qeneseq/qeneseq/AA1982.DAT.*
4: /SIDSB/qcdata/qeneseq/qeneseq/AA1983.DAT.*
5: /SIDSB/qcdata/qeneseq/qeneseq/AA1984.DAT.*
6: /SIDSB/qcdata/qeneseq/qeneseq/AA1985.DAT.*
7: /SIDSB/qcdata/qeneseq/qeneseq/AA1986.DAT.*
8: /SIDSB/qcdata/qeneseq/qeneseq/AA1987.DAT.*
9: /SIDSB/qcdata/qeneseq/qeneseq/AA1988.DAT.*
10: /SIDSB/qcdata/qeneseq/qeneseq/AA1989.DAT.*
11: /SIDSB/qcdata/qeneseq/qeneseq/AA1990.DAT.*
12: /SIDSB/qcdata/qeneseq/qeneseq/AA1991.DAT.*
13: /SIDSB/qcdata/qeneseq/qeneseq/AA1992.DAT.*
14: /SIDSB/qcdata/qeneseq/qeneseq/AA1993.DAT.*
15: /SIDSB/qcdata/qeneseq/qeneseq/AA1994.DAT.*
16: /SIDSB/qcdata/qeneseq/qeneseq/AA1995.DAT.*
17: /SIDSB/qcdata/qeneseq/qeneseq/AA1996.DAT.*
18: /SIDSB/qcdata/qeneseq/qeneseq/AA1997.DAT.*
19: /SIDSB/qcdata/qeneseq/qeneseq/AA1998.DAT.*
20: /SIDSB/qcdata/qeneseq/qeneseq/AA1999.DAT.*
21: /SIDSB/qcdata/qeneseq/qeneseq/AA2000.DAT.*
22: /SIDSB/qcdata/qeneseq/qeneseq/AA2001.DAT.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1660	93.5	634	22 AAM40125	Human polypeptide
2	1660	93.5	634	22 AAM92742	Human protein sequ
3	1660	93.5	634	22 AAM93478	Human protein sequ
4	1660	93.5	634	22 AAM95120	Human protein sequ
5	125	7.0	467	21 AAG58505	Arabidopsis thalia
6	103	5.8	291	21 AAY67420	Ursulin-gamma poly
7	102.5	5.8	4393	22 AAB31889	Amino acid sequenc
8	98	5.5	955	21 AAG30852	Arabidopsis thalia
9	98	5.5	1058	21 AAG30851	Arabidopsis thalia
10	98	5.5	1198	21 AAG30850	Arabidopsis thalia
11	93.5	5.3	400	22 AAG98362	Escherichia coli p

12	88.5	5.0	653	22 AAB47250	Human PP7. Homo s
13	88	5.0	1788	18 AAW23494	Murine Lysyl long
14	86.5	4.9	1002	21 AAG38741	Arabidopsis thalia
15	86	4.8	740	21 AAG42603	Arabidopsis thalia
16	86	4.8	790	21 AAG42602	Arabidopsis thalia
17	86	4.8	899	21 AAG42601	Arabidopsis thalia
18	85	4.8	510	20 AAY22201	Human extracellular
19	85	4.8	510	21 AAY54368	Protein encoded by
20	85	4.8	510	21 AAY66646	Membrane-bound pro
21	85	4.8	510	22 AAG03653	Human extracellular
22	85	4.8	510	22 AAB65169	Human PRO698 (UNO3
23	85	4.8	510	22 AAB50955	Human PRO698 (PRO6
24	85	4.8	2285	20 AAW98149	Bacillus subtilis
25	84	4.7	582	19 AAW62996	Human truncated St
26	84	4.7	701	14 AAR41335	84 kD ISGF-3alpha.
27	84	4.7	712	16 AAR72079	Human STAT4. Hom
28	84	4.7	712	17 AAW03170	Human STAT1-beta.
29	84	4.7	712	19 AAW62995	Human STAT1-beta p
30	84	4.7	739	14 AAR41334	91 kD ISGF-3alpha.
31	84	4.7	750	16 AAR72078	Human STAT1. Hom
32	84	4.7	750	17 AAW03168	Human STAT1-alpha.
33	84	4.7	750	19 AAW62994	Human STAT1-alpha
34	84	4.7	750	22 AAB19962	Human signal trans
35	84	4.7	786	21 AAD54098	Human pancreatic c
36	83.5	4.7	320	20 AAY34350	Porphyromonas qing
37	83.5	4.7	3801	18 AAW31949	Human b2 protein a
38	83	4.7	467	21 AAG23071	Arabidopsis thalia
39	83	4.7	508	19 AAW98852	H. pylori GHP1161
40	83	4.7	536	21 AAG31373	Arabidopsis thalia
41	83	4.7	537	21 AAG23070	Arabidopsis thalia
42	83	4.7	557	21 AAG23069	Arabidopsis thalia
43	83	4.7	606	21 AAG31372	Arabidopsis thalia
44	83	4.7	623	21 AAB25493	Eucalyptus grandis
45	83	4.7	626	21 AAG31371	Arabidopsis thalia

ALIGNMENTS

RESULT 1

AAM40125

ID AAM40125 standard; Protein: 634 AA.

XX AAM40125;

XX AAM40125;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 1270.

XX Human; non-tropic; immunosuppressant; cytostatic; gene therapy; cancer;
peripheral nervous system; neuropathy; central nervous system; CNS;
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chondrolytic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW Leukemia.

OS Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

PN W0200153312-AL.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000NOV-0534263.

XX 21-JAN-2000; 2000JUL-0534263.

XX 25-APR-2000; 2000JUL-0534263.

XX 09-JUL-2000; 2000JUL-0534263.

XX 19-JUL-2000; 2000JUL-0534263.

XX 03-AUG-2000; 2000JUL-0534263.

XX 14-SEP-2000; 2000JUL-0534263.

XX 19-OCT-2000; 2000JUL-0534263.

XX 29-NOV-2000; 2000JUL-0534263.

XX 29-NOV-2000; 2000JUL-0534263.

Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Glade, S., Gord, S., Goyette, M., Graham, L., Grand-pierre, N., Hados, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Katat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Melidrin, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Sougniez, C., Spencer, B., Stange-thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tophan, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (31-JUL-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 31, 2001 this sequence version replaced gi:14670134.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIDR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L18
Center clone name: L18_E_18

We are submitting the entire project as phase 3. The T7 end overlaps accession number AC005235 (WICGR project L389) by 18.8 kilobases.

FEATURES	source	Location/Qualifiers
	1. 166518	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="17"	
	/map="17"	
	/clone_lib="RPC1-11 Human Male BAC"	
repeat_region	49. 345	
repeat_region	/rpt_family="AluY"	
repeat_region	1704. 2064	
repeat_region	/rpt_family="L2"	
repeat_region	complement(2079. 2385)	
repeat_region	/rpt_family="AluX"	
repeat_region	2565. 2756	
repeat_region	/rpt_family="MIR"	
repeat_region	complement(3086. 3302)	
repeat_region	/rpt_family="AluSj/x"	
repeat_region	complement(3375. 3499)	
repeat_region	/rpt_family="MIR"	
repeat_region	3662. 3704	
repeat_region	/rpt_family="(GGGGA)n"	
repeat_region	4343. 4571	
repeat_region	/rpt_family="MIR"	
repeat_region	complement(4685. 4991)	
repeat_region	/rpt_family="AluSj"	
repeat_region	complement(6025. 6194)	
repeat_region	/rpt_family="MIR"	
repeat_region	complement(6198. 6413)	
repeat_region	/rpt_family="MER20"	
repeat_region	complement(6505. 6811)	
repeat_region	/rpt_family="AluX"	
repeat_region	7601. 7620	
repeat_region	/rpt_family="(CA)n"	
repeat_region	complement(8265. 8477)	
repeat_region	/rpt_family="MLT1J2"	
repeat_region	complement(8490. 8683)	
repeat_region	/rpt_family="MIR"	
repeat_region	8944. 8993	
repeat_region	/rpt_family="(TG)n"	
repeat_region	complement(9444. 9485)	
repeat_region	/note="<30 qual single clone coverage"	
repeat_region	complement(10463. 10532)	
repeat_region	/rpt_family="L3"	
repeat_region	complement(10696. 10829)	
repeat_region	/rpt_family="L2"	
repeat_region	complement(10942. 11071)	
repeat_region	/rpt_family="L2"	
repeat_region	complement(11559. 11717)	
repeat_region	/rpt_family="MIR"	
repeat_region	complement(11776. 12054)	
repeat_region	/rpt_family="AluSx"	
repeat_region	14340. 14492	
repeat_region	/rpt_family="MIR3"	
repeat_region	14999. 15060	
repeat_region	/rpt_family="MER41B"	
repeat_region	complement(15175. 15321)	
repeat_region	/rpt_family="MER5A"	
repeat_region	complement(15328. 15801)	
repeat_region	/rpt_family="LTR50"	
repeat_region	complement(15806. 15910)	
repeat_region	/rpt_family="MIR"	
repeat_region	16592. 16858	
repeat_region	/rpt_family="L2"	
repeat_region	17373. 17457	
repeat_region	/rpt_family="MER103"	
repeat_region	19279. 19576	
repeat_region	/rpt_family="L2"	
repeat_region	19778. 19830	
repeat_region	/rpt_family="L2"	
repeat_region	20330. 20626	
repeat_region	/rpt_family="AluSj"	
repeat_region	20857. 21030	
repeat_region	/rpt_family="MIR3"	
repeat_region	21878. 22012	
repeat_region	/rpt_family="MIR"	
repeat_region	complement(24397. 24685)	
repeat_region	/rpt_family="AluSj"	
repeat_region	complement(25649. 25738)	
repeat_region	/rpt_family="MIR"	
repeat_region	complement(25854. 25951)	
repeat_region	/rpt_family="MIR"	
repeat_region	complement(26035. 26101)	
repeat_region	/rpt_family="MIR"	
repeat_region	complement(26382. 26554)	
repeat_region	/rpt_family="MIR"	
repeat_region	complement(26567. 26957)	
repeat_region	/rpt_family="L2"	
repeat_region	27040. 27145	
repeat_region	/rpt_family="AluSx"	
repeat_region	complement(28771. 29065)	
repeat_region	/rpt_family="AluSj"	
repeat_region	29405. 29592	
repeat_region	/rpt_family="MER5A"	
repeat_region	complement(29769. 29891)	
repeat_region	/rpt_family="MIR3"	
repeat_region	complement(30555. 30773)	
repeat_region	/rpt_family="L1PA5"	
repeat_region	30883. 31182	
repeat_region	/rpt_family="AluSx"	
repeat_region	complement(31357. 31916)	
repeat_region	/rpt_family="MLT1G1"	
repeat_region	complement(31946. 32059)	
repeat_region	/rpt_family="MIR"	
repeat_region	32480. 32765	
repeat_region	/rpt_family="AluJo"	
repeat_region	34083. 34337	
repeat_region	/rpt_family="MIR"	
repeat_region	34392. 34426	

repeat_region	complement(3583, .3760)
repeat_region	/rpt_family="MER104"
repeat_region	complement(4048, .4165)
repeat_region	/rpt_family="L2"
repeat_region	4403, .4617
repeat_region	/rpt_family="GA-rich"
repeat_region	complement(4653, .4864)
repeat_region	/rpt_family="MER20"
repeat_region	complement(4958, .5158)
repeat_region	/rpt_family="MIR"
repeat_region	5311, .5586
repeat_region	/rpt_family="AluJo"
repeat_region	5727, .6106
repeat_region	/rpt_family="MUT21"
repeat_region	6825, .6855
repeat_region	/rpt_family="(TTTTTC)n"
repeat_region	complement(6858, .7141)
repeat_region	/rpt_family="AluSx"
repeat_region	complement(7502, .7793)
repeat_region	/rpt_family="AluY"
repeat_region	8799, .9098
repeat_region	/rpt_family="AluSq"
repeat_region	complement(9099, .9163)
repeat_region	/rpt_family="MER34B"
repeat_region	9342, .9382
repeat_region	/rpt_family="AT-rich"
repeat_region	9383, .9668
repeat_region	/rpt_family="AluSq"
repeat_region	9669, .9692
repeat_region	/rpt_family="AT-rich"
repeat_region	10135, .10331
repeat_region	/rpt_family="MIR"
repeat_region	10586, .10767
repeat_region	/rpt_family="MIR"
repeat_region	complement(10923, .11309)
repeat_region	/rpt_family="MUT1B"
repeat_region	11878, .12070
repeat_region	/rpt_family="MER33"
repeat_region	complement(12071, .12366)
repeat_region	/rpt_family="AluSx"
repeat_region	12367, .12466
repeat_region	/rpt_family="MER33"
repeat_region	complement(12511, .12731)
repeat_region	/rpt_family="MER20"
repeat_region	complement(12858, .12999)
repeat_region	/rpt_family="MIR"
repeat_region	13241, .13532
repeat_region	/rpt_family="AluSx"
repeat_region	complement(13554, .13670)
repeat_region	/rpt_family="L2"
repeat_region	13707, .13736
repeat_region	/rpt_family="AT-rich"
repeat_region	complement(14941, .15250)
repeat_region	/rpt_family="AluSn"
repeat_region	complement(15261, .15559)
repeat_region	/rpt_family="AluSq"
repeat_region	complement(15591, .15867)
repeat_region	/rpt_family="LIME"
repeat_region	complement(15884, .16024)
repeat_region	/rpt_family="MER102"
repeat_region	complement(16162, .16313)
repeat_region	/rpt_family="LIME/D"
repeat_region	16341, .16636
repeat_region	/rpt_family="AluSx"
repeat_region	16649, .16957
repeat_region	/rpt_family="AluY"
repeat_region	complement(17306, .17461)
repeat_region	/rpt_family="AluJo/FRAM"
repeat_region	complement(17465, .17758)
repeat_region	/rpt_family="AluSx"
repeat_region	complement(17857, .17998)
repeat_region	/rpt_family="MIR"
repeat_region	complement(18011, .18322)

[illegible]

Query Match	7.6%	Score 98.2	Dh 2	Length 190737
Best Local Similarity	85.8%	Prod. No. 9,4e+16		
Matches 109	Conservative 0	Mismatches 11	Indels 0	Gaps 0
Oy	943	tttccaatgatgagctctgcgtctcgtccccaagcctgaggaatcgaatgacagatctcgaac	1002	
Db 175584	TTTTGAGTATGAGAGTCTCGCTCTCTCCGCACTGCTGACGACGACGACGACGATCTTGCGC	175525		
Oy	1003	attgcagcctccaccctcctctgagttcaatgattcttccctgcttcaacctcccaagtaagctgga	1062	

Db	175524	ACTGACGCTCTACCTCCACGATTCCTGCTCAGCTCCACCTCCCAAGTACCTGGA	175465
Cy	1063	alttagc 1069	
Db	175464	ACTACAG	175458
RESULT	13		
LOCUS	AC025588		
DEFINITION	AC025588 145122 bp DNA	PM	12-MAR-2000
ACCESSION	AC025588		
VERSION	AC025588		
KEYWORDS	AC025588.1 GI:7229304		
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens		
REFERENCE	Baker, J. L., et al. (2000) Human Genome Project: The Human Genome Project. Nature 404: 456-457.		
AUTHORS	1 (bases 1 to 145122)		
TITLE	Lamerdin, J. E., McCreedy, P. M., Skowronski, E., Wisniewski, V.,		
JOURNAL	1 (bases 1 to 145122)		
COMMENT	Submitted (12-MAR-2000) DOE Joint Genome Institute, Lawrence		
	Livermore National Laboratory, Livermore, CA 94550		
	Map and sequence oriented from centromere to q-telomere. HMC		
	CIT-HSNC 46011 is overlapped on the 1c11 by HG249609,		
	(CIT-HSNC 258M23) and overlapped HG321036 (CIT-HSNC 444M24, AC005261)		
	on the 1c11 from bases 114293 to 145122 of this accession.		
	Additional chrom 14 map and sequence information are available at		
	http://www.himl.gov/bdtp/genome/genome.html .		
	Location/Qualifiers		
FEATURES	1..145122		
SOURCE	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="19"		
	/map="19q13.4"		
	/clone="CIT-HSNC 360011"		
	/clone_11b="CIT-HSNC"		
	/note="LLNL clone name: HC288669"		
	complement(185..562)		
	repeat_region		
	complement(775..931)		
	repeat_region		
	complement(1187..1486)		
	repeat_region		
	complement(1690..1831)		
	repeat_region		
	complement(1836..1960)		
	repeat_region		
	complement(2135..2253)		
	repeat_region		
	complement(2312..2405)		
	repeat_region		
	complement(2419..2494)		
	repeat_region		
	complement(3017..3126)		
	repeat_region		
	complement(3127..3179)		
	repeat_region		

```

* 83651 83750: gap of unknown length
* 83751 93283: contig of 9533 bp in length
* 93284 93383: gap of unknown length
* 93384 102351: contig of 8968 bp in length
* 102352 102451: gap of unknown length
* 102452 129254: contig of 26803 bp in length
* 129255 129354: gap of unknown length
* 129355 176697: contig of 47343 bp in length.
FEATURES             Location/Qualifiers
     source           1..176697
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /clone="RP11-655D7"
BASE COUNT  47001 a 41857 c 40829 g 44654 t  2356 others
ORIGIN
Query Match      7.6%; Score 98.2; DB 2; Length 176697;
Best Local Similarity 85.8%; Pred. No. 9.4e-16;
Matches 109; Conservative 0; Mismatches 18; Indels 0; Gaps 0:
QY 943 ttccaaatggagctcgttcgtctccagcagctgagttcaatgagcagctcgcgc 1002
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 175089 TTTTGAGATGGAGCTCGCTCTCTGCCCGAGCTGGAGTGCAGTGGCATCTTGGCTC 175030
QY 1003 attgcaagctccacctcagttcgaagcttctctcagcagctcctcagtagctagg 1062
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 175029 ACTGCAGCCTTACCTCCCGAGCTCAAGATCTCTCGCTCAGCTCCAGTAGTACGCGGA 174970
QY 1063 attacaq 1069
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174969 ACTACAG 174963

RESULT 12
AC019143/C
LOCUS          AC019143 190737 bp DNA             HFG      07-JUL-2000
DEFINITION    Homo sapiens chromosome 11 clone RP11-706A1}, WORKING DRAFT
SEQUENCE      28 unordered pieces.
AC019143
VERSION       AC019143.3 GI:7715065
KEYWORDS      HFG; HUGS_PHASE1; HUGS_DRAFT.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 190737)
AUTHORS      Waterston,R.H.
TITLE        The sequence of Homo sapiens clone
JOURNAL      Unpublished
REFERENCE     2 (bases 1 to 190737)
AUTHORS      Waterston,R.H.
TITLE        Direct Submission
JOURNAL      Submitted (30-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT       On May 7, 2000 this sequence version replaced gi:7024180.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_RH0706A1}
----- Summary Statistics -----
Sequencing vector: M13; 78%
Sequencing vector: plasmid; 22%
Chemistry: Dye-terminator ET; 78% of reads
Chemistry: Dye-terminator Big Dye; 22% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 167325 bases at least Q40
Consensus quality: 175317 bases at least Q30

```

FEATURES
source

Consensus quality: 180039 bases at least Q20
Insert size: 193000; avarose-fp
Insert size: 188037; sum-of-contigs
Quality coverage: 3.04 in Q20 bases; avarose-fp
Quality coverage: 3.18 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1325: contig of 1325 bp in length
* 1326 1425: gap of unknown length
* 1426 2586: contig of 1161 bp in length
* 2587 2687: gap of unknown length
* 2688 4338: contig of 1652 bp in length
* 4339 4438: gap of unknown length
* 4439 5989: contig of 1551 bp in length
* 5990 6090: gap of unknown length
* 6091 8270: contig of 2180 bp in length
* 8271 8370: gap of unknown length
* 8371 10239: contig of 1870 bp in length
* 10240 10339: gap of unknown length
* 10340 13484: contig of 3145 bp in length
* 13485 13584: gap of unknown length
* 13585 18186: contig of 4602 bp in length
* 18187 18286: gap of unknown length
* 18287 23764: contig of 5478 bp in length
* 23765 23864: gap of unknown length
* 23865 27992: contig of 4128 bp in length
* 27993 28092: gap of unknown length
* 28093 31755: contig of 3763 bp in length
* 31756 31855: gap of unknown length
* 31856 36229: contig of 4374 bp in length
* 36230 36329: gap of unknown length
* 36330 40157: contig of 3828 bp in length
* 40158 40257: gap of unknown length
* 40258 45494: contig of 5237 bp in length
* 45495 45594: gap of unknown length
* 45595 51920: contig of 6326 bp in length
* 51921 52020: gap of unknown length
* 52021 58426: contig of 6406 bp in length
* 58427 58526: gap of unknown length
* 58527 64892: contig of 6366 bp in length
* 64893 64992: gap of unknown length
* 64993 72930: contig of 7938 bp in length
* 72931 73030: gap of unknown length
* 73031 81807: contig of 8777 bp in length
* 81808 89406: gap of unknown length
* 89407 89506: gap of unknown length
* 89507 93183: contig of 9676 bp in length
* 93184 93283: gap of unknown length
* 93284 111814: contig of 12552 bp in length
* 111815 111934: gap of unknown length
* 111935 126259: contig of 14324 bp in length
* 126260 126358: gap of unknown length
* 126359 134920: contig of 8562 bp in length
* 134921 135020: gap of unknown length
* 135021 145335: contig of 10315 bp in length
* 145336 145435: gap of unknown length
* 145436 162083: contig of 16648 bp in length
* 162084 162183: gap of unknown length
* 162184 175617: contig of 13434 bp in length
* 175618 175717: gap of unknown length
* 175718 190737: contig of 15020 bp in length.

Location/Qualifiers
1..190737
/organism="Homo sapiens"
/db_xref="taxon:9606"

COMMENT

Direct Submission
Submitted (14-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis
MO 63108, USA
On Mar 13, 2000 this sequence version replaced gi:7024164.

NOTE: This is a working draft sequence. It currently consists of 1 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

/CLUSTER=KPL1 2200000
misc_feature
1., 196900

```

BASE COUNT	54333	a	41365	c	43067	g	58135	t
ORIGIN								

Query Match	7.6%;	Score 98.6;	Ds 2;	Length 196900;
Best Local Similarity	76.6%;	Pred. No. 7.3e-16;		
Matches 134; Conservative	0;	Mismatches 39;	Indels 2;	Gaps 1

901 cagcaltcagatqitaltclatcnaicvapuortqacvacatqetllecuaulqaaqlcteq 966

7588 CAGAGTTTATATTTTATTCITTCAGATACCATATTAATCTTTTTTTTCAGAGTGGAGCTCTTC 7647

967 attentatocceccacocatitcua!naccacaa!clcoaqclculliccauccleccaeclcc 1020

[illegible]

1698 C1C1G1CCCCA0U1U6W61UCAU1U6UCACU0A1C1CAU1CAG1U6CAWCC1CCCAC1C

1021 tgaattcaagc--ttctccctgcacagacccttcgaaatgatgtccttgc
||||| | |||||

7708 TGGGTTCAAGCAGTTCCTGCTCAAGCCTCCGGAGTAGCTGGGACTACAGGCAC 7762

07-7411-7000

	AC021170	176897 bp	DNA	MG	07-09-2000
DEFINITION	Homo sapiens chromosome 11 clone RP11-655D7, WOKING DRAFI				

SEQUENCE, 24 unordered pieces.
AC021170

```

VERSION          AC021170.4  C1:7235310
KEYWORDS
HTGS: HTGS PHASE1: HTGS DRAFT.

```

human.

ORGANISM	Phylogenetic position
<i>Homo sapiens</i>	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Primates; Hominidae; Homo
Eukaryota; Metazoa;	Eukaryota; Metazoa;

REFERENCE 1 (bases 1 to 176697)
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

AUTHORS	Waterston, R.H.
TITLE	The sequence of <i>Homo sapiens</i> clon-

Journal Unpublished
REFERENCE 2 (Pages 1 to 176697)

2 (cases 1 to 1000)
Walterston, K. H.

- * NOTE: This is a 'working draft' sequence. It currently
- * consists of 3 contigs. The true order of the pieces
- * is not known and their order in this sequence record is

----- Genomic Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://wugsc.wustl.edu/qsc/index.shtml>
----- Project Information -----
Center project name: H NH0255N19

Exon/Intron boundaries of identified genes were chosen, if there were canonical splice junctions that maintained sequence continuity across the splice junctions.

FEATURES

SOURCE

Location/Qualifiers

```

1..74155
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSH8-358P8"
/chromosome="X"
/clone_lib="Genome Systems Human BAC Library"
/map="Xp22"
1..206
/note="Overlap bases 137066 to 137270 in AC004552."
complement(282..561)
/rpl_family="Alusx"
878..1054
/rpl_family="MIR"
complement(1783..2080)
/rpl_family="Alu1b"
complement(2087..2340)
/rpl_family="Alusx"
complement(2384..2425)
/rpl_family="L2"
complement(2924..3108)
/rpl_family="L1MC4"
3138..3439
/rpl_family="Alusx"
3808..4265
/note="Region: Similar to M48206, Y22203, r1 Sources: fetal liver spleen INPLS Homo sapiens cDNA clone 243437 5'"
4150..4291
/rpl_family="ML17G"
complement(4703..4986)
/rpl_family="Alu1a"
5404..5576
/rpl_family="L1PA16"
complement(5425..5939)
/note="Region: N39012 Y22203.s1 Sources: fetal liver spleen INPLS Homo sapiens cDNA clone 243437 3'"
5987..6096
/rpl_family="Aluspy/4"
6479..6785
/rpl_family="Alusx"
7475..7776
/rpl_family="Alusx"
complement(8221..8347)
/rpl_family="FLAM_A"
complement(8989..9062)
/rpl_family="L1MC4"
9113..9370
/rpl_family="Alusx"
complement(9374..9745)
/rpl_family="MER31A"
9746..10021
/rpl_family="Alu1b"
complement(10022..10069)
/rpl_family="MER31A"
complement(10249..10269)
/rpl_family="AT-rich"
10297..10378
/rpl_family="MIR47H"
10379..10669
/rpl_family="Alusx"
10792..11085
/rpl_family="Alu1b"
complement(11793..11925)
/rpl_family="MIR"
join(12813..12955,16901..16972,21399..21583,23281..23370,37005..37082,44270..44346,45646..45747,46148..46258,49729..49842,52001..52497)
/genome="Homo sapiens pyruvate dehydrogenase kinase isoform 3"

```

```

repeat_region
/rpl_family="MER31B"
complement(14206..14403)
/rpl_family="L1MC4"
complement(14366..14460)
/rpl_family="(CA)n"
complement(14469..14724)
/rpl_family="L1MC4"
complement(14852..15151)
/rpl_family="Alu1"
complement(15355..15667)
/rpl_family="Alusx"
complement(17724..17814)
/rpl_family="MIR"
complement(18183..18506)
/rpl_family="MER2"
19504..19803
/rpl_family="Alusx"
20670..20933
/rpl_family="Alusq"
complement(24745..24860)
/rpl_family="L1MH6"
complement(24881..25180)
/rpl_family="Alusq"
complement(25183..25492)
/rpl_family="L1MH7"
25494..25640
/rpl_family="MIR"
complement(25752..25775)
/rpl_family="AT-rich"
complement(26089..26467)
/rpl_family="MSTA"
complement(26788..27088)
/rpl_family="Alusx"
27562..27637
/rpl_family="(CA)n"
29638..29936
/rpl_family="Alusq"
30636..30848
/rpl_family="Alusq/x"
complement(31901..31935)
/rpl_family="AT-rich"
32878..32956
/rpl_family="MER5A"
34392..34865
/rpl_family="L1MH7"
complement(34871..35166)
/rpl_family="Alusq"
35171..35227
/rpl_family="L1MH8"
complement(35295..35593)
/rpl_family="Alusx"
36273..36572
/rpl_family="Alusx"
46594..46775
/rpl_family="Alu1a"
36741..36982
/standard_name="dbSTS:47229, AFM189xc5, Chr. X, Homo sapiens"
36891..36954
/rpl_family="(CA)n"
37296..37428
/rpl_family="L1P5"
complement(37440..37741)
/rpl_family="Alusx"
37983..38280
/rpl_family="Alusx"
38284..38428
/rpl_family="L1MH"
39065..39370

```

	misc_feature	vp94d10.r1*	l4398..l4533	
Query Match	12.7%	Score 164.4:	DB 9: Length 154950:	
Best Local Similarity	96.6%:	Prct. No. 2,50-11:		
Matches 168:	Conservative 0:	Mismatches 6:	Indels 0:	Gaps 0:
OY 1122	tcaatcttccacagatcatcgtattttggaaggagaccacatttgcatcttcnaaa	1181		
Db 6844	TTCACATCTCTTCAAGATTCACGACTATTTCGACAGCAGCATCATATGTGGCATTTTGAACA	6903		
OY 1182	aagaagctctcacaacalcccacaccgcagcgcaaatctggqctcataaacgccaanact	1241		
Db 6904	AAAAGCTCGTCACACATFACGCATCCGACCAAAATTCAGCTCATFTAACGCCACAGAAGCT	6963		
OY 1242	atttcagctcttttaaaattcattcagccccaatatgtcataatttcgcgaagaagtcc	1295		
Db 6964	ATTTCAGCTCTCTTAACCTTCATTTGCCCTFAAATGTCATTCGTGAGATATTTTC	7017		
RESULT 8				
LOCUS AC004656	74155 bp	DNA	PRT	02-JUL-1998
DEFINITION Homo sapiens Xp22 HAC GSHH-35BPH (Genome Systems Human BAC Library)				
ACCESSION AC004656				
VERSION complete sequence.				
KEYWORDS AC004656.1 GI:3253114				
SOURCE HTG.				
ORGANISM human.				
REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:				
AUTHORS Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.				
1 (bases 1 to 74155)				
Muzny,D., Akenson,A.D., Adams,C., Brandage,E., Bunae,G.C.,				
Carvelli,K., Chacko,J., Chen,Y., Di,W., Ding,Y., Duang,S.,				
Durbin,J., Foreman,J., Ganesh,R., Garcia,C., Goodman,M.,				
Corrall,J.H., Haywood,M., Hernandez,J., Jackson,L., Jin,S.,				
Kampal,I.R., Karpathy,S., Kovar,C., Leal,B., Li,Y., Liechfield,O.,				
Liu,W., Logan,O., Lu,J., Ly,T., Martinez,C., Oswal,G., Perez,L.,				
Rashid,N.D., Rowland,K., Savage,L., Scherpf,S.E., Shen,H.,				
Simon,M., Stovall,K., Tims,K.M., Todd,J., Vo,O., Williamson,A.,				
Worley,K.C., Yu,W., Chinault,C., Nelson,D. and Gibbs,R.A.				
JOURNAL Direct Submission				
REFERENCE Unpublished				
AUTHORS 2 (bases 1 to 74155)				
JOURNAL Worley,K.C.				
TITLE Direct Submission				
JOURNAL Submitted (02-MAY-1998) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
REFERENCE 3 (bases 1 to 74155)				
AUTHORS Worley,K.C.				
JOURNAL Direct Submission				
REFERENCE Submitted (25-JUN-1998) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
AUTHORS 4 (bases 1 to 74155)				
JOURNAL Morley,K.C.				
REFERENCE Direct Submission				
AUTHORS Submitted (02-JUL-1998) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
COMMENT On Jun 25, 1998 This sequence version replaced gi:3242666.				
The sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality does not meet this standard, it will be indicated in the annotation.				
The repeat regions shown were identified using RepeatMasker by Adrian Smith.				
Sequence similarities were identified using Powerblast by Jinghui Zhang.				

TITLE	Shibahara,T., Tanaka,T. and Nakamura,T.							
JOURNAL	Direct Submission Submitted (29-AUG-2000) to the DDBJ/EMBL/Genbank databases. Sunio Sunguo. Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center; Shitokame-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:sda@mls.u-tokyo.ac.jp, Tel:+81-3-5449-5286, Fax:+81-3-5449-5416)							
COMMENT	MEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction: Y's & T'-and one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).							
FEATURES								
source	location/Qualifiers 1..1752 /organism="Homo sapiens" /db_xref="taxon:9606" /cclone="HS102573" /clone_lib="HS1" /tissue="tyc="human small intestine" /note="cloning vector pMEBsf3" 135..1289 /note="unnamed protein product" /codon_start=1 /protein_id="BAB15A04.1" /db_xref="GI:10439020" /translat ion="MNSNOMYMSLEFHLGCTSYGRIRVPENSDOLKDKDES ANITAMQSHDVTQSDEPDSPSSGSVLVEVERCOILATLTTHQYLIERLPICATI KOKAQRYRIAKLRKYRPRLFPGSVKLMPKHLLQEYPHEDDDIIFODARKTPP VOIAWTSYDSKKTWTKNQGRKVAVHFVNKNLILPMSECLILECGLSELKSLS KRNSTIIPARSGHELLEDLDISAPFLIGTTHNGCKCCSLSIIONLSLVMTSMILH SSVAALQIVPLDYVFYWFRTFDGTGVLEVAYIMDSDFEQHASPVMLMDQLQSVCS MMNDPEPCIKIDVYFMVKEFKFSYNWTNKTNONTCYOIIIPPYVANHYT"							
BASIC : L ⁵⁵⁵	A	555	A	419	C	414 G 544 T		
ORIGIN								
Query Match:	32.1%	Score	416.2;	DH	9;	Length	1752;	
Basic Local Similarity	81.8%;	Pred.	No. 3.5e+100;					
Matches	599;	Conservative	0;	Mismatches	3;	Indels	121; gaps	2
OY	616	GATTCAAGTACAACTCATCGACTCACAAAACCTGGACAAGAAGAACCTTAGGCACGAATAAC	675					
Db	1	GATTTAAGTCACATGCCATGGCTTACCAAAC - CTCACAAATAGACATTTTAACTTGACATGAAT	59					
OY	676	GAATGTCATATGGAGAAGATCTCTTAAGTCTGGAAGCTTCTTAAAGATGCTATAGCTCCA	735					
Db	60	GATTTTCATTTTCGCAAGATCTCGAAGAGCTTGCAACGTTTTCTTAATCTATACGCTTCA	119					
OY	736	ACCAGAACTCAATCAATCAATTCAGAAATCGAAGAACCTGACATCTAGATCTCTCATCTCT	795					
Db	120	ATCGGACATTCATCATGANTGANTTCAAGACATGACAAATGCTTAAGCTTAACTTGCATCTT	179					
OY	796	CATGAATGAATCAATCTACAGCTCTGAGATCTGAGATCTCTGCGAGAAAAATCAATCTCAATG	855					
Db	180	CATTCAGAGTACCGAGTTACGCTCGCGCAATCACAGCTTCCTGCCAGAAAGTACATCTGATCG	239					
OY	856	AGTCAATCTGAAAAGAGATCTTAAATCTGAAATCTGAAGATCAATCAATCAATCAATCAAT	915					
Db	240	GATTAACATCTGAANAAGCATTTAAAGATCTTGAAATTTGACAGCGCAATCAGCATTCAGATCT	299					
OY	916	ATCTGTCAATCAATCAATCAATCAATCAATCTGCAAAATTAAGTCTGAGCTCTGCTCTCAAG	975					
Db	300	ATCTGTCAATCAATCAATCAATCAATCAATCTGCAAAATTAAGTCTGAGCTCTGCTCTCAAG	344					
OY	976	AGTCAATCTGAAAAGAGATCTTAAATCTGAAATCTGAAGATCAATCAATCAATCAATCAAT	1035					
Db	334	-----	333					
OY	1036	CTGGCTCTCAAGCTCTCCAAATTAATGAATCAATCAATCAATCAATCAATCAATCAATCAAT	1095					

DB	1388	-----	GCCTGGATCA	1398
QY	1081	gtatcatctaaagatgaagaatgctcaaaagcatctgctcaaatattctcaatgc	1140	
DB	1399	gtatcatctaaagatgaagaatgctcaaaagcatctgctcaaatattctcaatgc	1458	
QY	1141	caatctctggaaggagaccacacatgctgcatctctgaacaagaagctctcaaacac	1200	
DB	1459	caatctctggaaggagaccacacatgctgcatctctgaacaagaagctctcaaacac	1518	
QY	1201	ggcaccggagcaaaattgaatgcatataaagccaaaagatctcttccatctgttaac	1260	
DB	1519	ggcaccggagcaaaattgaatgcatataaagccaaaagatctcttccatctgttaac	1578	
QY	1261	catgcccctaaatgcatctctgctcaaaagctccaca	1298	
DB	1579	catgcccctaaatgcatctctgctcaaaagctccaca	1616	
RESULT	4			
LOCUS	AK001230			
DEFINITION	Homo sapiens CDNA FLJ10368 fls, clone NT2RM2001544.	PRI	22-FEB-2000	
ACCESSION	AK001230			
VERSION	AK001230.1			
KEYWORDS	GI:7022354			
SOURCE	oligo capping; fls (full insert sequence);			
ORGANISM	Homo sapiens testicular carcinoma cell_line:NT2 CDNA to mRNA,			
	clone_lib:NT2RM2 clone:NT2RM2001544.			
	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
REFERENCE	1 (sites)			
AUTHORS	Isoqali,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,			
	Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,			
	Magatsuma,M., Hosofuji,T., Kaku,Y., Kohira,H., Kondo,H.,			
	Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,			
	Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,			
	Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,			
	Nakamura,Y., Nagahara,K., Masuho,Y., Ninomiya,K. and Iwyanagi,T.			
	NEDO human cDNA sequencing project.			
	Unpublished (2000)			
	2 (bases 1 to 2611)			
TITLE	Isoqali,T. and Otsuki,T.			
AUTHORS	Direct Submission			
JOURNAL	Submitted (16-FEB-2000) to the DDBJ/EMBL/Genbank databases, Tokyo			
	Isoqali, Helix Research Institute, Genomics Laboratory, 1532-1 Yama,			
	Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,			
	Tel:81-448-92-4951, Fax:81-448-92-4952)			
	NEDO human cDNA sequencing project supported by Ministry of			
	International Trade and Industry of Japan; cDNA full insert			
	sequencing: Research Association for Biotechnology; cDNA library			
	construction, 5'- & 3'- and one pass sequencing and clone selection;			
	Helix Research Institute (supported by Japan Key Technology Center			
	etc.) and Department of Virology, Institute of Medical Science,			
	University of Tokyo.			
COMMENT	Location/Qualifiers			
	1..2611			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/cell_line="NT2"			
	/cell_type="testicular carcinoma"			
	/clone="NF2RM2001544"			
	/clone_lib="NT2RM2"			
	/note="cloning vector: pME18SFL3; mRNA from uninduced NT2			
	neonatal precursor cells."			
	24..1928			
	/note="unnamed protein product"			
	/odon_start=1			
	/protein_id="BA01568.1"			
	/db_xref="GI:7022354"			
	/translation="MSLVATPV LTPMLDVKCTLVNNGGVYKRPPLSKTQYV			
	SVYTFVDQVVKITCLLPSNTNVALP LTKRNDI VPHIRIK LQYKKEKQGLTSSGFA			
	SLTEFDGLADPLIPRTSKYFNFTEDHKWVLELRWASVTHMSVMTLLKLDVQPMQ			

Query Match	Best Local Similarity	80.74%	Score 1048:	DH 9:	Length 2611:
Matches 1178:	Conservative	0:	Mismatches	0:	Indels 120: Gaps 1:
QY	1	algccttcggttcgagcaacaaatcaatataatacaaccccaatacaactcaaggtatg	60		
DB	24	atgcctcttggttcgcagcaacaaatvatvatatacaccccttaacaaacttaagcctgct	83		
QY	61	acaactctcaatgctctabagatattcaaaagttctcttcaagccccaatactcaagaaaqa	120		
DB	84	acaattctcaaatgctctatgctcttttttcgaactttttaaagccccaattatcagcaaaaga	143		
QY	121	actaatctatctgacatctcaactcaatctatgagccagaaatctgaanaactcaactctgcg	180		
DB	144	actgatatttgcctcagttgtatcttatcttgagaccagcaaatgttaacttaactgcctcg	203		
QY	181	cgcctctatctggaactatagaaagccctctcaataactctcaaaaagagagatctgtctgc	240		
DB	204	cctctttagcgaactatcgaaccccttcccaattatttaaaaatgagagatttctgtgcg	263		
QY	241	ttctcaagagctgaanaactcaaaatcaataaaagagagacatcaggtatcaacaaactctgc	300		
DB	264	ttttcacagctgaacatcaaaatcaaaatvatvatataaaaagcagactcaagcagactctgcg	323		
QY	301	ttctgacactctcaagctctcgaaggaactcttggaagccctatactatcactcgcgaactcaagc	360		
DB	324	ttttgcatactttttgaactttttgaagaaattttttgaagccctcttatctatctgacacttgcagc	383		
QY	361	aagaaattcttactctcaactctgaagccagcaaaaatctgacaaagctctctgactctcaagca	420		
DB	384	aagtaatttttaacctttgacatctgacagcagccagcaaaaatgctgacagccttgcgtttgcgca	443		
QY	421	tcgaatctcaatctcaagctctctgaagaattcaataaattctgagatctctgaagcaatcagc	480		
DB	444	ttctaatcttatctgacactctgaagccagcaaaaatctgacaaagctctctgactctcaagca	503		
QY	481	taactctgaagcctcaactctgacagactctctgaggaagcaaaaatctgagagctgacactctct	540		
DB	504	tatttttgacactgactcttgcagctctttttgacgaagaatcaaaagctgacagaaatctgattctgt	563		
QY	541	ctaaagagatctgaagctgagcaacagagacactctgactctctgagagttctcaatacaaac	600		
DB	564	ctaaagactatgacagctgacagcagcaacaaacaaattttgacagctgacagctgacagctgac	623		
QY	601	cttctctctctgaagagatcaattctgaatctgactctgactctgactctgactctgactctgact	660		
DB	624	cttctctctctgacagctgactgattgaatctgacatctgacatctgacatctgacatctgacatct	683		
QY	661	ctaaagctctgaagcaagctctgactctgacagagactctgagagctctgagagctctctctctga	720		
DB	684	tttaactctgacagatgaatgactgactgactgactgactgactgactgactgactgactgactgac	743		
QY	721	actaatatgactctcaatacaactctcaatcaatcaatctcaagagatctgacagatctcaat	780		
DB	744	atctaatatgactctcaatacaactctcaatcaatcaatctgacagatctgacagatctcaat	803		
QY	741	ctaaagctctgactctcaagagctgacagatctgagctgagagactctgagagctctgagag	840		
DB	804	tttaagctctgactctcaagagctgacagatctgacagatctgacagatctgacagatctgacag	863		

DE HEPARAN SULFATE PROTEOGLYCAN PERLECAN (FRAGMENT).
GN HSPG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
PN [1]
RP SEQUENCE FROM N.A.
RA Genoscope;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RS SEQUENCE FROM N.A.
RX MEDLINE=20553141; PubMed=11101850;
RA Nicole S., Davoine C.S., Topaloglu H., Cattolico L., Barral D.,
RA Beighton P., Ben-Hamida C., Hammouda H., Cruaud C., White P.S.,
RA Samson D., Urtizberea J.A., Lehmann-Horn F., Weissenbach J.
RA Hentati F., Fontaine B.
RI "perlecan, the major proteoglycan of basement membranes, is altered in
RI patients with Schwartz-Jampel syndrome (chondrodystrophic myotonia)."
RL Nat. Genet. 26:480-483(2000).
CC -I- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
CC (LDLRA) DOMAIN.
CC EMBL; AL445795; CAC18534.1; -;
DR InterPro: IPR000551; EGF-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR003599; Ig_2.
DR InterPro: IPR003598; Ig_c2.
DR InterPro: IPR003600; Ig-like.
DR InterPro: IPR003596; Ig_v.
DR InterPro: IPR000034; Laminin_B.
DR InterPro: IPR002049; Laminin_EGF.
DR InterPro: IPR001791; Laminin_G.
DR InterPro: IPR002172; LDL_recept_A.
DR InterPro: IPR000082; SEA.
DR Pfam: PF00052; laminin_B_3.
DR Pfam: PF00053; laminin_EGF_8.
DR Pfam: PF00054; laminin_G_3.
DR Pfam: PF00057; ldl_recept_a; 4.
DR Pfam: PF01390; SEA; 1.
DR PRINTS: PR00261; LDLRECEPTOR.
DR PRODOM: PD003031; Laminin_B; 3.
DR SMART: SM00181; EGF; 15.
DR SMART: SM00180; EGF_Lam; 12.
DR SMART: SM00001; EGF-like; 7.
DR SMART: SM00409; IG; 22.
DR SMART: SM00408; IGc2; 22.
DR SMART: SM00406; IG; 7.
DR SMART: SM00410; IG-like; 1.
DR SMART: SM00281; LamB; 3.
DR SMART: SM00282; LamG; 3.
DR SMART: SM00192; LDla; 4.
DR SMART: SM00200; SEA; 1.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS50068; LDLRA_2; 4.
KW Glycoprotein.
FT NON_TER
SQ SEQUENCE 4370 AA; 466592 MW; F8D762C3940A557C CRC64;

Query Match 5.8%; Score 102.5; DB 4; Length 4370;
Best Local Similarity 24.1%; Pred. No. 19;
Matches 84; Conservative 39; Mismatches 140; Indels 95; Gaps 14;

Qy 48 TIVDTNFKLTCILFSGNYEALPIIKNGDI-----VREHLKI-----GVYKKE *2
Db 2234 TVAEGOTLDLSCVY-AGAAHQAQVWYRGSGSLPARHQVGRSLYIFQASPADAGQVVCRA 2292
Qy 93 TOGITSSGFASLIEFGLGA-----PI-IPRTSSKYFNFIED----- 129
Db 2293 SNGMEAS--IIVTGTGAGANLAYPAGSTOPRIEPPSSQVAEQITLDLNCVVFQSHAO 2350
Qy 130 ---HMVEALRWASTHMSPSWTLKLCQVQPMQYFDLTCLLGKAEVDCGASFLKKVWDG 136
FT NON_TER
GN

Db 2351 VTNHKGSGSLPVRHQTHGS-----LLRLYQASPADSGEYVCRVLGSSVPLEASVLTIEPA 2406
Qy 187 TRIP-----FPSRVLIQDLVLEGLDLSHRLQNLITDILVYDN-HVHV-----AR 231
Db 2407 GSVFALGVIPIVRI-----ESSSQVAEQITLDLNCVLAGQAQVTHKRGGSILPAR 2459
Qy 232 SLKVGSGFLRI-----YSLHTKLOSNNSENQTMLSLEFHLHGGSYGRGIRVLPS 281
Db 2450 HQVHGSRLRLLOVTPADSGEYVCRVVGSGGTQEAASVLTIOORLSGSHSGVAYPVRIES 2519
Qy 282 NSDVDQLKKDLESANLHANQHSQVICOSEPDSPN-----GVSLRPP 324
Db 2520 SSASLANGHILDLNCLVASOAPHIITWYKRGSLPSRHOIVGSRLRIP 2567
RESULT 13
Q9XWB2 PRELIMINARY; PRT; 215 AA.
AC Q9XWB2:
DI 01-NOV-1999 (TRENBLrel. 12, Created)
DI 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DI 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
DE 3RS.1 PROTEIN.
GN 3RS.1.
CS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Barlow K.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL02671; CAA21777.2; -;
DR EMBL; 281568; CAA21777.2; JOINED.
DR EMBL; 281568; CAB76729.1; -;
DR EMBL; AL032671; CAB76729.1; JOINED.
SQ SEQUENCE 215 AA; 25209 MW; BE4975C081625390 CRC64;

Query Match 5.7%; Score 100.5; DB 5; Length 215;
Best Local Similarity 25.9%; Pred. No. 0.43;
Matches 48; Conservative 25; Mismatches 81; Indels 31; Gaps 7;

Qy 160 QYFDLTCLLGKAEVDGASFLKLVWDGTR-----TPFSPRVLIQDLV 202
Db 26 QYFDLLAQVHVSVEITMDGLMRLVWRAQKFGPESIKERREROLFHVTOFSFKRYIVPPNP 85
Qy 203 LEGDLSHRLQNLITDILVYDNHVVHVARSLKVGSLFYSLTKLQSNNSNQTMLSLE 262
Db 86 RIGKAIEFGKEVL-IEINVDYDEHRADLYSLNSGNFVAIONVBA-ASTPHREIQI----- 138
Qy 263 FHLH-GGTSYGRGIRVLPSNSDVS---OLKKDLESANLHANQHSQVICOSEPDSPN 318
Db 139 --LHGGEAYFQGISIVP-VDFEVDFAQNFKRKVESVLENVLYDENFIEFOQPEEVIENR 195
Qy 319 VSLRP 323
Db 196 VPEEP 200
RESULT 14
Q04538 PRELIMINARY; PRT; 1198 AA.
ID Q04538:
DI 01-JUL-1997 (TRENBLrel. 04, Created)
DI 01-JUL-1997 (TRENBLrel. 04, Last sequence update)
DI 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE SIMILAR TO SYNECHOCYSTIS ANTIVIRAL PROTEIN.
GN F20P5.20.

```

Qy 112 APII-----PRTSKYFNFTIEDHKMVEAL-----RWASTHM 141
D 133 IPVYPASAOVALLPFGKOGESHKLKEEHQVSYLYNWIDKYDVPDEAEYQOORVKSUNV 192
Qy 145 SPSNTLLKCDVQVDFDLTCOL-----LGKAEV-----DGASFLKLVYDGTGTRIP 190
D 193 KDRFSLLK--DVTGNYFDLQGVAKDPYDEMGRIIYLSYDENTDLFFHYHTEGVRDL 250
Qy 191 FPSNR---VLIODLVLEGDLSH---IHLQNLITIDILYDNNHVVHVAR--SLKVGSLFLRYS 243
D 251 ASAAPADAYDPEDNNPEAAQHPWGPYGRKTIQISCTYDAHADFIKAGVSAGNWLSLRN 310
Qy 244 LHTK 247
D 311 VOVK 314

RESULT 10
Q93702 PRELIMINARY: PRI: 587 AA.
AC Q93702
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DI 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE F39H2.1 PROTEIN.
GN F39H2.1
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Kershaw J.;
RL Submitted (0CT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Haynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen P.,
RA Smalldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston P.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.
RL Nature 358:32-38(1994).
DR EMBL: Z81080; CAB03086.1;
SQ SEQUENCE 587 AA: 55602 MW: CF8A55F14188446F CRC64;

```

```

Query Match 5.18; Score 109; DB 5; Length 587;
Best Local Similarity 22.98; Pred. No. 0.31; Mismatches 60; Indels 89; Gaps 10;
Matches 51; Conservative 24;

Qy 122 YNFTIEDHKMVE-----ALRWASTHMSPSWILLKLCVQPMQYFDLI-----COL-- 158
D 71 YFDVLAQVHSYVETKNGSWILLRWRAQKFGPEASKEK----RKNLPHVIENTFKCYIVP 126
Qy 169 -----LGKA-EVDGASFLKLVNDGTGTRIPPSNRVLIODLVLEGDLSHILRLONLITIDILVY 22?
D 127 PDPRIGRAVEESGREFL-----IEIDVY 149

Qy 224 DNNHVVHVARSLKVGSLFLRYSLSH-----TKLOSMNSNOTMLSLFHLHGTSYGRGIRV 277
D 150 DDPHIGIEKLSGDFVALONVHAASVGLIEMQVHLG-----GGQATNRGISK 196

Qy 278 LPES--NSDVOLKLDLSA-----NLTANSHDVICOSE 310
D 197 VPVDFRNEAFOIFKRKRVSVESDAVTDGDNFIEFOCKENNVQSE 239

```

```

RESULT 11
Q9V4E4 PRELIMINARY: PRI: 1044 AA.
ID Q9V4E4
AC Q9V4E4
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DI 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE KIAA0580 PROTEIN (FRAGMENT).
GN KIAA0580.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
PC TIISUE=BRAIN;
RX MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N.,
RA Ohara O.;
RI "Prediction of the coding sequences of unidentified human genes. IX.
RI The complete sequences of 100 new cDNA clones from brain which can
RI code for large proteins in vitro.";
RL DNA Res. 5:31-39(1998).
DR ENBL: AB011152; BAA25506.1;
DR InterPro: IPR000198; RhoGAP.
DR InterPro: IPR001164; Znf_GCS.
DR InterPro: IPR001849; PH.
DR Pfam: PF00169; PH; 4.
DR Pfam: PF00620; RhoGAP; 1.
DR Pfam: PF01412; ArfGAP; 1.
DR PRINTS: PR00405; REVINTRACING.
DR PROSITE: PS50003; PH_DOMAIN; 3.
DR SMART: SM00105; ArfGAP; 1.
DR SMART: SM00233; PH; 3.
DR SMART: SM00324; RhoGAP; 1.
DR NON_TER
FT NON_TER
SQ SEQUENCE 1044 AA: 119453 MW: 68E22A6DEAA53BD CRC64;

Query Match 5.88; Score 103; DB 4; Length 1044;
Best Local Similarity 24.08; Pred. No. 2.3;
Matches 70; Conservative 39; Mismatches 119; Indels 64; Gaps 15;

Qy 49 IVDSQVNTKLTCL---FSGNYEALPIIKNGDIVRHLKIQVYKKTQGTSSGFSALT 105
D 255 LLEETN-KKMCVLEGGFLSYENDKSTTPNGTININEVICIAHKEDFYLTNGPIF---I 310
Qy 106 FEGTLGAPIIPRTSSKYFNFTIEDHKMVEALRWVA---STHMSPSWILLKLCVQPMQYF 162
D 311 FE-----IYLPSEVFLFGAETS---QAQRKWTETAIKHFVPLFA-ENLEAD----Y 355
Qy 163 DLTCOLLCKAEVDGASFLKLVW---DGTTRTPPSWRVLIQDLVLEGLDLSHILRLONLITID 219
D 356 DLICQLFYK-DCHALDOWRKGNFAMDKSSLHF-----CLMQEYQGDGMHLRLLOELTIS 409
Qy 220 ILVYDNN-VHVARSLKVGSLFLRYSLSHTKL-----QSMNSNOT 257
D 410 IMVNGEKELDVLLVLEKGRITLYHG-HTKLDFTVWHTAIEKAAAGTDGNALODQOLSKNDV 468
Qy 258 MSLSEFHLHGTSYGRGIRVLPESND-----VDOLKLDLESANLTANOH 302
D 469 PIIVNSCIAFTVQGLGCKYIYQKNGDPLHISELSEFFKDKDARSKFLRAGKH 520

```

```

RESULT 12
Q9H3V5 PRELIMINARY: PRI: 4370 AA.
ID Q9H3V5
AC Q9H3V5
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DI 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

```


DB 2351 VTHKRGSLPVRHQTGSGS---LLRLVOASPADSGEYVCRVLGSSVPLEASVLVTIEPA 2406
QY 187 TRIP---FPSNRVLIDQLVLEGDLSHIRLQNLITDILVYDN-HVHV-----AR 231
DB 2407 GSVFALGVITPVI-----ESSSSQVAEGOTLDLNCVAGQAQVTHKRGSLPAR 2459
QY 232 SLKVGSGFLPI-----YSLHTKLQSMNSNOTMLSLEFHLHGTSYGRGIRVLPS 281
DB 2460 HQVHGSRLRLQVTFADSGEYVCRVVGSGGTOEASVLVTIQORLSGSHSQGVAPVRIES 2519
QY 282 NSVDVLDLKKDLKESANLTANQHSVDICQSEPDOSFPN-----GVSLRPP 324
DB 2520 SSASLANGHLDLNCVLAQAPHIITWYKRGSLPSRHQIVGSRRLRIP 2567
RESULT 13
ID Q9XNB2 PRELIMINARY: PRT: 215 AA.
AC Q9XNB2:
DI 01-NOV-1999 (TrEMBLrel. 12, Created)
DI 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DI 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE 3R5.1 PROTEIN.
GN 3R5.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae.
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=5239;
RN [1]
RP SEQUENCE FROM N.A.
PA Barlow K.;
RL Submitted (OCI-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA McMurra; A.A.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL032671; CAA21777.2;
DR EMBL; Z81568; CAA21777.2; JOINED.
DR EMBL; Z81568; CAB76729.1;
DR EMBL; AL032671; CAB76729.1; JOINED.
SO SEQUENCE 215 AA: 25209 MW: BE4975C081625390 CRC64;

Query Match 5.7% Score 100.5; DB 5; Length 215;
Best Local Similarity 25.9%; Pred. No. 0.43; Indels 31; Gaps 7;
Matches 45; Conservative 25; Mismatches 81;
QY 160 QYFDLTCOLLGKAEVDGASFLKVMGCTR-----TPFSSNRVLIDQLV 202
DB 26 QYFDLLAQVHVSVMETMDGLWHLRVRAQKFGPESIKERRERQLEHVTQFSFKRIVPPNP 85
QY 203 LEGULSHIRLONLITDILVYDNHVVARSILKVGSLRIYSLTKLQSMNSNOTMLSLE 262
DB 86 RIGKAIEEFGKEYL-IEINVDYDEHRAVLVSLNSGNFVAIQNVHA-ASTPHREIQI----- 138
QY 263 FHLH--GGTISYGRGIRVLPSNDVD---OLKKDLKESANLTANQHSVDICQSEPDOSFPNG 318
DB 139 --LHGGEAYORGISTVP-VDFVDVAFQNFRRKRVESLVLENVLYDENFIEFOOPEEVIENR 195
QY 319 VSLRP 323
DB 196 VPREP 200
RESULT 14
ID Q04538 PRELIMINARY: PRT: 1198 AA.
AC Q04538:
DI 01-JUL-1997 (TrEMBLrel. 04, Created)
DI 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DI 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE SIMILAR TO SYNECHOCYSTIS ANTIVIRAL PROTEIN.
GN F20PE.24.

DE HEPARAN SULFATE PROTEOGLYCAN PERLECAN (FRAGMENT).
GN HSPG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Genoscopy;
RL Submitted (OCI-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
EX MEDLINE=20553141; PubMed=1101850;
RA Nicole S., Devoline C.S., Topaloglu H., Cattolico L., Barral D.,
RA Beighton P., Ben-Hamida C., Hammouda H., Cruaud C., White P.S.,
RA Samson D., Urtizberea J.A., Lehmann-Horn F., Weissenbach J.,
RA Hentati F., Fontaine B.;
RI patients with Schwartz-Jampel syndrome (chondrodystrophic myotonia).;
RL Nat. Genet. 26:480-483(2000).
CC -!- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
(LDLRA) DOMAIN
EMBL; AL445795; GAC18534.1;
DE InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR003599; Ig.
DR InterPro: IPR003598; Ig_c2.
DR InterPro: IPR003600; Ig-like.
DR InterPro: IPR003596; Ig_Y.
DR InterPro: IPR000034; Laminin_B.
DR InterPro: IPR002049; Laminin_EGF.
DR InterPro: IPR001791; Laminin_G.
DR InterPro: IPR002172; LDL_recept_A.
DR InterPro: IPR000082; SEA.
DR Pfam: PF00052; laminin_B; 3.
DR Pfam: PF00053; laminin_EGF; 8.
DR Pfam: PF00054; laminin_G; 3.
DR Pfam: PF00057; ldl_recept_a; 4.
DR Pfam: PF01390; SEA; 1.
DR PRINTS: PR00261; LDLRECEPTOR.
DR ProDom: PD003031; Laminin_B; 3.
DR SMART: SM00181; EGF; 15.
DR SMART: SM00180; EGF_Lam; 12.
DR SMART: SM00001; EGF_Like; 7.
DR SMART: SM00409; IG_22.
DR SMART: SM00408; IG_c2; 22.
DR SMART: SM00406; IGV; 7.
DR SMART: SM00410; IG_Like; 1.
DR SMART: SM00281; Lamb; 3.
DR SMART: SM00282; Lamb; 3.
DR SMART: SM00192; LDLa; 4.
DR SMART: SM00200; SEA; 1.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS50058; LDLFA_2; 4.
KW Glycoprotein.
FT NON_TER
SO SEQUENCE 4370 AA: 465592 MW: F8D762C3940A557C CRC64;

Query Match 5.9% Score 102.5; DB 4; Length 4370;
Best Local Similarity 24.1%; Pred. No. 19; Indels 85; Gaps 14;
Matches 84; Conservative 39; Mismatches 140;
QY 48 TIVDTNVKLTCLLFGNVEALPIIYKNGDI-----VFHRLKI-----QYKKE 92
DB 2234 TVAGOTLDLSCV-AGQAQVTHKRGSLPARHQVFGSRLYFQASPADAGQVYCR 2292
QY 93 TOGITSSGFASLTGEGILGA-----PI-IPRTSSKYFNFTED----- 129
DB 2293 SNGMEAS--IIVTVTQGANLAYPAGSTOPRIEPSQSSQVAEGOTLDLNCVVGSHQ 2350
QY 130 ---HKWFAALRWASTIHSPSTILLKCDQPMQYFDLTCOLLGKAEVDGASFLKVMG 185

```

|||||
Db 1211 CGCATCCGAGCAAAATTGAGTGCATATTAAGCCCAAGAGACTATTTCAGTCTGTAAACTT 1270
|||||
Qy 1261 cattgccctaatactcattgctgaagaagtccaca 1298
|||||
Db 1271 CATTGCCCTAAATGCTATTGCTGCAAGAAGTTCACCA 1308
|||||

RESULT 3
LOCUS AK022580 2383 bp mRNA PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA FLJ12518 fis, clone NT2RM2001805.
ACCESSION AK022580
VERSION AK022580.1 GI:10434050
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens teratocarcinoma cell_line:NT2 cDNA to mRNA.
clone_lib:NT2RM2 clone:NT2RM2001805.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y.,
Matsumawa,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,
Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Sasaki,N.
NEDO human cDNA sequencing project
Unpublished (2000)
2 (bases 1 to 2383)
Isogai,T. and Otsuki,T.
Direct Submission
Submitted (23-AUG-2000) to the DDBJ/EMBL/GenBank databases. Takao
Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp,
Tel:81-438-52-3951, Fax:81-438-52-3952)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing and clone selection;
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of virology, Institute of Medical Science,
University of Tokyo.
FEATURES
Location/Qualifiers
1..2383
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone_lib="NT2RM2001805"
/clone="NT2RM2"
/notes="cloning vector: pME18SFL3-mRNA from uninduced NT2
neuronal precursor cells."
439..2343
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAB14110.1"
/db_xref="GI:10434051"
/translation="MSLVPTNYYITPLNOLKGTIVNYYGVKFFKPPVLSKGDYIC
SVTVIDQTVKPTCLLFCSGNYEALPIIYKNGDIVREHLKIOYVKKETOGITSSGPA
SLIEFGILNAPLIPRTSSKYNFTEDHKWEALRVWASHMSPSWTLKLCVOPNQ
YFDLTCQLLGAEDVASFLKVDWDTPTPPSWRLIQDLVLEGLDLSHLRLONLTI
DILYNDHVHVARSLKVGSEFLRIYSLHTKLOSMNSENOTMLSEFLHLGGTSGYGRIR
VLPSNDPOLKTKLESLANLTANQSDVICOSEPDSDSFTSSGSVSLYEYERCOOLSA
TILTDHOYLRPLCALILKQKAPQOYIRAKLRYSKPRFLQSVKLCPRKHLIQEVP
HEGDLIIIFDQATKTPDVKLONTSLYDSKIWTITKNQGRKVAHVFKNNGLPLSNE
CLLIIEGGLTSEICKLSNKTNSVIPRSGHEDLELLDLSAPFLIOGTIRHHYCKOCS
LRSIQNSLSNLDKTSWIPSSVAEALGIYVPLYVFMFTFLDDGTGVLVLAIMSDKEF
QIPASEVIMDDLDQKSDVMIMDMFCPPGKIDATPWLDECIFIKSTNVINGIDNQICYOI
EDTTVAEDVI"
BASE COUNT 750 a 456 c 459 g 718 t
ORIGIN

```

Query Match 80.7%; Score 1048; DB 9; Length 2383;

```

Best Local Similarity 90.8%; Pred. No. 7.5e-268;
Matches 1178; Conservative 0; Mismatches 0; Indels 120; Gaps 1;
Qy 1 atgtctttggttcagcaacaattatataacacccctggaatcaacttaagggtggt 60
|||||
Db 439 ATGCTTTGGTTCAGCAACAATAATATATATACACCCCTGAATCAACTTAAGGTGT 498
|||||
Qy 61 acaattgtcaatgtctatggtgttggaagtctttaaagcccccatatcacaagaagga 120
|||||
Db 499 ACAATTGTCAATGTCTATGTTGTGTGAAGTCTTTTAAGCCCCCATATCTAAGCAAGGA 558
|||||
Qy 121 actgattattgctcagttgtaactattgtgaccagacacaaatgtaaaactaaacttgctg 180
|||||
Db 559 ACTGATTATTGCTCAGTTGTAAGTATTGTGACCAGACAAATGTAACAACTTAACCTG 618
|||||
Qy 181 ctctttagtgaactatgaagccctcccaataattataaaatggagagatatgttctgc 240
|||||
Db 619 CTCTTTAGTGAACACTATGAAGCCCTTCCCAATAATTATATAAATGGAGATATTGTTCGC 678
|||||
Qy 241 ttccacaggtgaagattcaagtatataaaaggagactcaggttatccacagctctg 300
|||||
Db 679 TTTCACAGGCTGAAGATTCAGGTATATAAAGAGAGACTCAGGGTATACACAGCTCTGGC 738
|||||
Qy 301 ttgtcatctttgacgtttgagggaactttggagccctcatatcatcgcacacttcaagc 360
|||||
Db 739 TTTGCATCTTTGACGTTTGAGGAACTTTGGAGCCCTATCATACCTCGACCTTCAAGC 798
|||||
Qy 361 aagtattttaactcactactgagaccacacaaatgtagaagccttacgtgtttgggca 420
|||||
Db 799 AAGTATTTTAACTTCACTACTGAGGACCACANAATGTTAGAACCTTACGTGTGTTGGCA 858
|||||
Qy 421 tctaactatgtccacgctttggacattactaaaaattgtgtgatgttcagccaatgcag 480
|||||
Db 859 TCTACTCATATGTCAACGCTCTTGGACATTACTAAAAATGTGTATGTTCAGCCCAATGCAG 918
|||||
Qy 481 tatttgaactcacttcagctcttggaagcaagcaagtgacgagcagcatcttctt 540
|||||
Db 919 TATTTTACGCTGACTTGTGACGCTTTGGCAAGCAAGTGGACGGAGCATCATTTCTT 978
|||||
Qy 541 ctaaaagtgatgggagggccacagacaccatttccattcttggagagtccttaatacaagc 600
|||||
Db 979 CTAAAGGTATGGGATGGCAGCAGGACACCACTTTCCATCTTTGGAGAGTCTTAAATCAAGAC 1038
|||||
Qy 601 ctgttcttgaagtgatttaagtccatccatcgcgtcacaatactgacaatagacatt 660
|||||
Db 1039 CTGTGTTCTTGAAGTGATTTAAGTCACATCCATCGGCTACAAAATCTGACAAATAGACATT 1098
|||||
Qy 661 ttagtctacataaccatgttctatgtggcaagatctctgaaggttggaagctttcttaga 720
|||||
Db 1099 TTAGTCTACGATAACCATGTTTCATGTGGCAAGATCTCTGAAGGTTTGAAGCTTTCTTTAGA 1158
|||||
Qy 721 atctatagccttcataccaaacttcaatcaatgaattcagagaatcagacaatgttaagt 780
|||||
Db 1159 ATCTATAGCCTTTCATACCAAACTTCAATCAATGAATTCAGAGAAATCAGACAAATGTTAAGT 1218
|||||
Qy 781 tttagagtttcatcttcattgaggttaccagtttacggttcggggaatcagggtttgtccagaa 840
|||||
Db 1219 TTAGAGTTTTCATCTTCATGGAGGTACCAAGTTACGTCGGGGAATCAGGGTCTTGCCAGAA 1278
|||||
Qy 841 agtaactctgatggtggaactcaactaaaaggatttaatactcgaatttgacagccaat 900
|||||
Db 1279 AGTAACCTCTGAIGTGGATCAACTGAAAAGGATTTAGAATCTGCAAAATTTGACAGCCAAT 1338
|||||
Qy 901 cagcattcagatgttatctgtcaatcagaacctgacagcagcttcccaaatggagctcg 960
|||||
Db 1339 CAGCATTACAGATGTTATCTGTCAATCAGAACCTTGACGACAGCTTTTCCAA----- 1387
|||||
Qy 961 ctctgctccagctggaggttcagtgagcaggtctcggctcattgcagcctccacctcc 1020
|||||
Db 1388 ----- 1387
|||||
Qy 1021 tgagttcaagctctctcctgctcagctcccaagtagctgggattacaggctctggatca 1080
|||||

```


Source: *Humanities*
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

[illegible][illegible]

```
Query Match      20.5%; Score 390.8; DB 11; Length 546;
Best local Similarity 99.5%; Pred. No. 9,5e-84;
Matches 392; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1506 atgtaaacagtgatctctgagatccatccaaatctaaattcccttgatctgataaaac 1565
    |||
Db 1 atgtaaacagtgatctctgagatccatccaaatctaaattcccttgatctgataaaac 60

Oy 1566 atcgtggattccctctctgagagagagcactgggtatgtgacccctccaatatgigt 1625
    |||
Db 61 atcgtggattccctctctgagagagagcactgggtatgtgacccctccaatatgigt 120

Oy 1626 tgttatgaccttaccacitgatgagagagaggtactagaaagccttatctcatggattc 1685
    |||
Db 121 tctttatgacctttacacittgatgatggaacagagagactactagaagccttatctcatggattc 180

Oy 1686 tgaacaattcttccagatccagaculcagaggttctgagtgatgataaccttcaagaaag 1745
    |||
Db 181 tgaacaattcttccagatccagaculcagaggttctgagtgatgataaccttcaagaaag 240

Oy 1746 tgtggatatgatcgtggatatgtttgtccctccaggaataaaatgataatataaccgtg 1805
    |||
Db 241 tctggatatgatcgtggatatgtttgtccctccaggaataaaatgataatataaccgtg 300

Oy 1806 attgaatgcttcatcgaattctatcacatgtccacaaatgagacaaataaatcaaatctacta 1865
    |||
Db 301 gtggaaatgctttcatcgaattctatcacatgtccacaaatgagacaaataaatcaaatctacta 360

Oy 1866 tcaaatlltttgaacacacagatctcagagagatata 1899
    |||
Db 361 tcagatTTTTGACACACACAGTTGCAGAGAGATCAA 394
```

Search completed: April 9, 2002, 22:13:52
Job time: 8867 sec

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	112	19.8	493	1	TEBA_STJMY	P29550	stylychnia
2	104.5	18.4	446	1	TEB_EUPCR	Q06184	eploies cr
3	97	17.1	495	1	TEBA_OXyno	P29549	oxytricha n
4	90	15.9	460	1	TEBH_EUPCR	Q05183	eploies cr
5	72.5	12.8	1073	1	HSER_PIG	P55204	sus scrofa
6	70.5	12.4	286	1	R1P1_CUCFI	Q619x4	cucumis fig
7	69.5	12.3	1161	1	DP3A_AQUAE	Q67125	aquifex aeg
8	68	12.0	1324	1	COPA_BOVIN	P27954	bos taurus
9	67	11.8	87	1	SVR_SALIN	P74871	salmonella
10	67	11.8	1256	1	MRP_STFSU	P32653	streptococc
11	65.5	11.7	801	1	RV3A_SCHPO	P05020	schizosacch
12	66	11.6	336	1	RUVB_HELPJ	Q08557	helicobacte
13	66	11.6	336	1	RUVB_HELPY	Q25699	helicobacte
14	66	11.6	1251	1	RBP2_PLAVB	Q00799	plasmodium
15	65	11.5	221	1	G12Z_ARATH	Q92xv3	arabidopsis
16	65	11.5	577	1	PRIM_BUCAL	P57164	buchnera ap
17	65	11.5	2244	1	PPY1_SCHPO	Q00794	schizosacch
18	64.5	11.4	412	1	FKR4_SPOFR	Q25486	spodoptera
19	64.5	11.4	577	1	SYR_HAEIN	P43832	haemophilus
20	63.5	11.2	122	1	Y01Z_NFVAC	P41422	autographa
21	63	11.1	433	1	SECY_RICPR	Q92cs5	rickettsia
22	63	11.1	1224	1	COPA_HUMAN	P53621	homo sapien
23	62.5	11.0	1073	1	HSER_HUMAN	P25092	homo sapien
24	62.5	11.0	3082	1	POLG_ZYHVS	Q36979	z genome po
25	62	10.9	282	1	YKCO_CAEEL	P42001	caenorhaudiv
26	62	10.9	396	1	DDL_TREPA	Q83676	treponema p
27	62	10.9	421	1	TYPH_MYCGE	P47297	mycoplasma p
28	62	10.9	524	1	YMOH_YEAST	P54074	saccharomyc
29	62	10.9	877	1	DP01_STRPN	P13252	streptococc
30	62	10.9	986	1	EP1B_STAEP	P30195	staphylococ
31	62	10.9	1450	1	MP5F_CHICK	Q02173	gallus gall
32	61.5	10.8	442	1	ETBR_RAT	P21451	ratus norv
33	61.5	10.8	928	1	NRPI_XENLA	P28924	xenopus lae

Q: 52 OTNVKLTCLLFSGNYEALPIIYKNGDIVRFHRLKIQVYKKETO 94

00: 52 0TNVKITCILESGNYFAPIIYKNGDIVREHRL

管


```

Db 115 RAGDIIRVHRATFLYNGORO 135
      : |||:| || :||:| :|
RESULT 4
TEBH_EUPCR STANDARD: PRT: 460 AA.
AC Q06183;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE TELOMERE-BINDING PROTEIN HOMOLOG.
OS Euplotides crassus.
OC Eukaryota: Alveolata: Ciliophora: hypotrichs: Euplotida: Euplotidae;
OC Euplotes.
OC NCBI_TaxID=5936;
RN [1]
PP SEQUENCE FROM N.A.
RX MEDLINE-93126105; PubMed-1460483;
RA Wang W., Stopp R., Scofield M., Price C.;
PI telomere-binding protein homologs.;
PL Nucleic Acids Res. 20:5621-5629(1992).
CC -!- FUNCTION: MAY BIND TELOMERIC TAG4 SEQUENCES.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE TELOMERE-BINDING PROTEIN ALPHA SUBUNIT
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M96819; AAA29128.1;
CC PIR: S35525; S35525.
CC InterPro: IPR003415; Telo_bind_alpha.
CC Pfam: PF02307; Telo_bind_alpha; 1.
CC DNA-binding, Nuclear protein; Telomere; Multigene family.
CC KW SEQUENCE 460 AA: 53360 MW: EDIC141385A0B5FE CRC64:
CC -----
Query Match 15.9% Score 90: DB 1: Length 450:
Best Local Similarity 31.4% Pred. No. 0.022:
Matches 33: Conservative 12: Mismatches 46: Indels 14: Gaps 4:
Oy 9 YIYPLNOL--KGGIIVNYGVVKKFFRPPLSKGIDYCSVVIIVDQI-NVK-----LIC 59
Db 19 YEYIEGSIIEENEFASINFAVVIDACFPYKVDKKYKCYLKVIDITHNVKRGDDREAIIV 78
Oy 60 LLESGNVEALPIYKNGDIVRFHELKI-----QVYKKEIOGITS 29
Db 79 ALQSRKFDLPIORCGDIIRVHRAEYNYKDDQHYFKLNYSYSS 123
RESULT 5
HSEF_PIG STANDARD: PRT: 1073 AA.
AC P55204; Q29050;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HEAT-STABLE ENTEROTOXIN RECEPTOR PRECURSOR (GC-C) (INTESTINAL
DE GUANYLATE CYCLASE) (EC 4.6.1.2) (5TA RECEPTOR).
GN GUCY2C OR GUC2C.
OS Sus scrofa (Pig).
OC Eukaryota: Metazoa: Chordata: Graniata: Vertebrata: Euteleostomi;
OC Mammalia: Eutheria: Cetartiodactyla: Suina: Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
PP SEQUENCE FROM N.A.

```

```

EC TISSUE=Small intestine;
RX MEDLINE-95058376; PubMed=7969686;
RA Wada A., Hirayama I., Kitao S., Fujisawa J.-I., Hidaka Y.,
RA Shimonishi Y.;
PI pig intestinal membrane-bound receptor (guanylyl cyclase) for heat-
PI stable enterotoxin: cDNA cloning, functional expression, and
PI characterization.;
RL Microbiol. Immunol. 38:535-541(1994).
RN [2]
RN REVISIONS TO 238 AND 509.
RA Wada A.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: RECEPTOR FOR THE E-COLI HEAT-STABLE ENTEROTOXIN (E.COLI
CC ENTEROTOXIN MARKEDLY STIMULATES THE ACCUMULATION OF CGMP IN
CC HAMMALIAN CELLS EXPRESSING GC-C). ALSO ACTIVATED BY THE ENDOGENOUS
CC PEPTIDE GUANYLIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: GTP = 3', 5'-CYCLIC GMP + PYROPHOSPHATE.
CC -!- SUBCELLULAR LOCATION: TYPE 1 MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-4/GUANYLYL CYCLASE
CC FAMILY.
CC -!- SIMILARITY: CONTAINS 1 PROTEIN KINASE-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D17513; BAA04455.1;
CC HSP: Q02846; IAWL.
CC InterPro: IPR001828; ANF_receptor.
CC InterPro: IPR000719; Euk_pkinase.
CC InterPro: IPR001054; Guanylt_cyclase.
CC Pfam: PF01094; ANF_receptor; 1.
CC Pfam: PF00211; guanylate_cyc; 1.
CC Pfam: PF00069; pkinase; 1.
CC SMART: SM00044; CYCC; 1.
CC PROSITE: PS00452; GUANYLATE_CYCLASES_1; 1.
CC PROSITE: PS00125; GUANYLATE_CYCLASES_2; 1.
CC PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
CC Receptor; Transmembrane; Glycoprotein; Lyase; CGMP synthesis; Signal.
KW SIGNAL 1 23
FT CHAIN 1 24 1073 HEAT-STABLE ENTEROTOXIN RECEPTOR.
FT DOMAIN 24 433 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 434 454 POTENTIAL.
FT DOMAIN 455 1073 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 489 749 PROTEIN KINASE-LIKE.
FT DOMAIN 824 954 GUANYLATE CYCLASE.
FT CARBHYD 32 32 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBHYD 43 43 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBHYD 79 79 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBHYD 195 195 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBHYD 224 284 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBHYD 307 307 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBHYD 357 357 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBHYD 402 402 N-LINKED (GLCNAC...) (POTENTIAL).
FT SEQUENCE 1073 AA: 123219 MW: 5644FC1327BA9F43 CRC64:

```

```

Query Match 12.8% Score 72.5: DB 1: Length 1073:
Best Local Similarity 29.5% Pred. No. 4.7:
Matches 43: Conservative 19: Mismatches 39: Indels 21: Gaps 7:
Oy 14 LNOIKGGIIV---NHYGVVKKFFRPPLSKGT---DYCS---VVTIVDQHYFKLTLFSG 64
Db 255 VHILKGGRAVEDIIVLDLFDNHYFDNVTPADYKKNVLVLTPPENSYSNS--FSK 312
Oy 65 NVEAL----PIYKNGDIVRFHRLKIQVYKKEIOGITS 108
Db 312 ILSLVKNDIFLAYHNGVLLFQHMLKIFLEKRE--DVITSKFAHAFRNITEG 362

```

```

RESULT 6
RIP1_CUCFI
ID RIP1_CUCFI STANDARD: PRI: 286 AA.
AC Q9FRX4;
DI 20-AUG-2001 (Rel. 40, Created)
DI 20-AUG-2001 (Rel. 40, Last sequence update)
DI 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE RIBOSOME-INACTIVATING PROTEIN PRECURSOR (RRNA N-GLYCOSIDASE)
DE (EC 3.2.2.22).
OS Cucumis figareli.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Cucurbitales; Cucurbitaceae; Cucumis.
OX NCBI_TaxID=131071;
EN [1]
PP SEQUENCE FROM N.A.
RA Yamada T., Ohki S.T., Osaki T.;
RI "Cloning and analysis of a cDNA coding a putative ribosome-
PI inactivating protein from Cucumis figareli.";
PL Plant Biotechnol. 17:337-340(2000).
CC -!- CATALYTIC ACTIVITY: ENDOLYOLYSIS OF THE N-GLYCOSIDIC BOND AT ONE
CC -!- SPECIFIC ADENOSINE ON THE 28S RPNA.
CC -!- SIMILARITY: TO OTHER BACTERIAL AND PLANTS RIBOSOME-INACTIVATING
CC PROTEINS. BELONGS TO TYPE 1 RIP.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AB045560; BAB19677.1;
CC InterPro: IPR001574; RIP.
CC Pfam: PF00161; RIP; 1.
CC PRINTS; PR00396; SHIGARICIN.
CC PROSITE; PS00275; SHIGA_RICIN; 1.
KW Antiviral; Protein synthesis inhibitor; Hydrolase; Toxin; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 286 POTATIVE RIBOSOME-INACTIVATING PEPTIDE.
FT ACT_SITE 185 185 BY SIMILARITY.
FT CARBOHYD 103 103 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 110 110 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 252 252 N-LINKED (GLCNAC...) (POTENTIAL).
FT SEQUENCE 286 AA; 31771 MW; 4EFD4966E604DA41 CRC64;
SQ
Query Match 12.4%; Score 70.5; DE 1; Length 286;
Best Local Similarity 25.6%; Pred. No. 1.7;
Matches 30; Conservative 13; Mismatches 37; Indels 37; Gaps 5;
QY 3 LVPATN-----YIVPLNOLKGGIIVNVYGVVKKFKPPYLSKGTIDYGSVVTIVDQ 59
||||: || ||: || ||: || ||: || ||: || ||: || ||: || ||: || ||: ||
Db 63 LVPISGSPBYILMOLSNVEGRTITMAVDVTRVYHGVYVNGISY-----FFMEINELIA 117
QY 59 -----CLLFSONEALPIIYKNSDIVRHLKIQVYKKEKGTGITSQFAS; 194
||||: || ||: || ||: || ||: || ||: || ||: || ||: || ||: || ||: ||
Db 118 SKVFOGKSIILPYSONYOKLO-----SVAPKERDSI-FLCFMA; 157
RESULT 7
DP3A_AQUAE
ID DP3A_AQUAE STANDARD: PRI: 1161 AA.
AC Q67125;
DI 15-DEC-1998 (Rel. 37, Created)
DI 15-DEC-1998 (Rel. 37, Last sequence update)
DI 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA POLYMERASE III ALPHA SUBUNIT (EC 2.7.7.7).
DN DNAAE OR AQ.1009.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.

```

```

OX NCBI_TaxID=633563;
RH [1]
PC SEQUENCE FROM N.A.
RF STPAIN-VF5;
RX MEDLINE=98195656; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RI "The complete genome of the hyperthermophilic bacterium Aquifex
RI aeolicus.";
RL Nature 392:353-358(1998).
CC -!- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
CC RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
CC THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
CC THE ALPHA CHAIN IS THE DNA POLYMERASE (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =
CC N PYROPHOSPHATE + DNA(N).
CC -!- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND THETA
CC CHAINS) THAT ASSOCIATES WITH A TAU SUBUNIT WHICH ALLOW THE CORE
CC DIMERIZATION TO FORM THE POLIII' COMPLEX. POLIII' ASSOCIATES WITH
CC THE GAMMA COMPLEX (COMPOSED OF CHAINS GAMMA, DELTA, PSI,
CC AND CHI) AND WITH THE BETA CHAIN (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-C FAMILY. DNAAE
CC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE000718; AAC07087.1;
CC InterPro: IPR003141; PHE-N.
CC DR Pfam: IPR002309; TRNA-synt_2.
CC DR Pfam: PF02231; PHE-N; 1.
CC DR Pfam: PF01335; TRNA-enti; 1.
CC DR SMART: SM00481; POLIIIAC; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication;
KW Complete proteome.
SQ SEQUENCE 1161 AA; 133207 MW; 619F7436E1262BD4 CRC64;
Query Match 12.3%; Score 69.5; DB 1; Length 1161;
Best Local Similarity 29.3%; Pred. No. 11;
Matches 22; Conservative 12; Mismatches 24; Indels 17; Gaps 4;
QY 5 PATNY-----IYPLNLK-----GGTIVNVYGVVKKFKPPYLSKGTIDYGSVVTIVDQ 52
||||: || ||: || ||: || ||: || ||: || ||: || ||: || ||: || ||: ||
Db 972 PLDNVEKLLKNRYTPIEDLEENDKESEAVLTGVITELKVKK----TKNGDYMAVFNLDK 1027
QY 53 TNVKTCLLESQNE 67
||||: || ||: || ||: || ||: || ||: || ||: || ||: || ||: || ||: ||
Db 1028 TGLIECVFPGVYE 1041
RESULT 8
CPFA_BOVIN
ID CPFA_BOVIN STANDARD: PRI: 1224 AA.
AC Q27554;
DI 01-MCV-1997 (Rel. 35, Created)
DI 01-MCV-1997 (Rel. 35, Last sequence update)
DI 20-AUG-2001 (Rel. 40, Last annotation update)
DE CALICMODER ALPHA SUBUNIT (ALPHA-COAT PROTEIN) (ALPHA-COP) (HEPCOP)
DE (HEP-COP) (CONTAINS: XENIN (XENOPSIN-RELATED PEPTIDE); PROXENIN).
GH CPA.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;

```

[illegible]

```

RI Streptococcus suis type 2.*;
RL Infect. Immun. 60:2351-2367(1992).
CC -!- SUBCELLULAR LOCATION: TYPE 1 MEMBRANE PROTEIN. CELI WALL.
CC -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X64450; CAA45781.1;
DR PIR: A43829; A43822.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00745; Gram_pos_anchor; 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
KW Repeat: Transmembrane; Cell wall; Signal.
FT SIGNAL 1 47
FT CHAIN 48 1255 MURAMIDASE-RELEASED PROTEIN.
FT DOMAIN 48 1231 EXTRACELLULAR (POTENTIAL).
FT TRANSHEM 1232 1248 MEMBRANE ANCHOR.
FT DOMAIN 1249 1255 CYTOPLASMIC (POTENTIAL).
FT REPEAT 563 681 SMALL.
FT REPEAT 839 861 SMALL.
FT REPEAT 953 1005 LAPCE.
FT REPEAT 1064 1084 SMALL.
FT REPEAT 1089 1142 LARGE.
FT REPEAT 1143 1195 LARGE.
FT DOMAIN 1223 1228 CONSERVED IN GRAM-POSITIVE CELL SURFACE
SO SEQUENCE 1256 AA: 135795 MW: DCF7F5242F14341 CRC64:

Query Match 11.8% Score 67: DB 1: Length 1256:
Best Local Similarity 29.2% Pred. No. 22:
Matches 28: Conservative 15; Mismatches 33; Indels 2: Gaps 5:

OY 7 TNYIYPLNOLKGTIVNVY-----GVVKKFKP-----PYLSKGTGYCSVTIVDQITKKL 57
DB 1000 TTHYKTIPEYKNGIVNVNVTEDGVTK--EPVDTPTSGETPY-----DILGGRK 1050
OY 58 TLLFSG-NYEALPII-YKMGDIVRHLKIOVYKK 91
DB 1051 KIITFKGEYELVRVDGIENCKVVEGETIVTVYVYRK 1086

RESULT 11
VP34_SCHPO
ID VP34_SCHPO STANDARD: PPT: 801 AA.
AC P50520; Q3P3W3; Q9UPD2.
DI 01-OCT-1996 (Pel. 34, Created)
DI 20-AUG-2001 (Pel. 40, Last sequence update)
DI 20-AUG-2001 (Pel. 40, Last annotation update)
DE PHOSPHATIDYLINOSITOL 3-KINASE VPS34 (EC 2.7.1.137) (PI3-KINASE)
DE (PTDINS-3-KINASE) (PI3K) (VACUOLAR SORTING PROTEIN 34).
GN VPS34 OR SPAC458.05.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID:4896;
[1]
PN PN
PP SEQUENCE FROM N.A.
FX MEDLINE=96350252; Pubmed=8719881;
PA Takegawa K., Dewald D.B., Emr S.E.;
FT Schizosaccharomycetes pombe vps34p, a phosphatidylinositol-specific PI
FT 3-kinase essential for normal cell growth and vacuole morphology.
PL J. Cell Sci. 105:3745-3756(1995).
PN [2]
PP SEQUENCE FROM N.A.

```

```

PC STEAIN-972;
PA Harris D., Lyne M., Rajandream M.A., Barrell B.G.;
RN Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
PP SEQUENCE OF 138-801 FROM N.A.
PX MEDLINE=95290763; Pubmed=7772832;
PA Kimura K., Miyake S., Makuuchi M., Morita R., Usui T., Yoshida M.,
PA Harinouchi S., Fukui Y.;
FT Phosphatidylinositol 3-kinase in fission yeast: a possible role in
FT stress responses.
RL Biosci. Biotechnol. Biochem. 59:678-682(1995).
CC -!- FUNCTION: PHOSPHATIDYLINOSITOL 3-KINASE HOMOLOG REQUIRED FOR
CC VACUOLAR SORTING AND SEGREGATION
CC -!- CATALYTIC ACTIVITY: ATP + 1-PHOSPHATIDYL-ID-MYO-INOSITOL = ADP +
CC 1-PHOSPHATIDYL-ID-MYO-INOSITOL 3-PHOSPHATE.
CC -!- SIMILARITY: BELONGS TO THE PI3/P14-KINASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U92593; AAC19133.1;
DR EMBL: AL358272; CAB93947.1;
DR InterPro: IPR002420; PI3K_C2.
DR InterPro: IPR001263; PI3Ka.
DR InterPro: IPR000403; PI3_P14_kinase.
DR Pfam: PF00613; PI3Ka; 1.
DR Pfam: PF00792; PI3K_C2; 1.
DR Pfam: PF0454; PI3_P14_kinase; 1.
DR SMART: SM00145; PI3Ka; 1.
DR SMART: SM00146; PI3K; 1.
DR SMART: SM00142; PI3K_C2; 1.
DR PROSITE: PS00915; PI3_4_KINASE_1; 1.
DR PROSITE: PS00916; PI3_4_KINASE_2; 1.
DR PROSITE: PS00290; PI3_4_KINASE_3; 1.
KN Transferase; Kinase.
FT DOMAIN 541 799 PI3K/PI4K.
FT CONFLICT 164 164 K -> O (IN REF. 1).
FT CONFLICT 236 236 L -> I (IN REF. 3).
FT CONFLICT 542 542 S -> T (IN REF. 3).
SQ SEQUENCE 801 AA: 92135 MW: 520571E1475CC341 CRC64:

Query Match 11.7% Score 66.5: DB 1: Length 801:
Best Local Similarity 29.9% Pred. No. 15:
Matches 3: Conservative 9; Mismatches 24; Indels 21; Gaps 3:

OY 23 VVVYGVVKKFKPKPYLSKGTGYCSVTIVDQITKKLFLFSGNVEALPIIYKNGDIVRPH 82
DB 509 VRIYVRII-----PDAC---TVFKTHQPLRLFLFKCGSKYPIIIFKNGDRLRQD 554
OY 93 RIKYIV-----YKKE 92
DB 555 CVVQILLMDKLIKKE 571

RESULT 12
RUVF_HELPJ
ID RUVF_HELPJ STANDARD: PPT: 336 AA.
AC Q9ZM57.
DI 30-MAY-2000 (Pel. 39, Created)
DI 30-MAY-2000 (Pel. 39, Last sequence update)
DI 20-AUG-2001 (Pel. 40, Last annotation update)
DE HOLLAND JUNCTION DNA HELICASE RUVF.
GN RUVF_P_RHP0365.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID:85963;

```


Very Match	11.5%	Score 65:	DB 1:	Length 221:
Best Local Similarity	21.1%	pred. No. 5.1:		
Best Local Similarity	19.0	Matches	33:	Models 54:
Best Local Similarity	29	Concentrating		Gaps 5:

C:Genetics:
A:Gene: dnaE
C:Superfamily: DNA-directed DNA polymerase III alpha chain

Query Match 12.3%; Score 69.5; DB 2; Length 1161;
Best Local Similarity 29.3%; Pred. No. 34;
Matches 22; Conservative 12; Mismatches 24; Indels 17; Gaps 4;

Oy 5 PATHV-----IYFLNQLK-----GGTIVNVGVVKFFKPYPYLSGTDYCYSVTIDQ 52
||| | : : : || : || : || :
Db 972 PLDNYELKNRYPIEDLEWDKESEAVLTGVTILKKVK---TKNGDYMAVENLVDK 1027

Oy 53 TNVLCTLLFSGNVE 57
| : : || : ||
Db 1028 TGLTECVWFGGYE 1041

RESULT 13
E75338
DNA gyrase, subunit A - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: E75338
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
M.: Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.: Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286: 1571-1577, 1999
A:title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: E75338
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-812 <WHI>
A:Cross-references: GB:AE002030; GB:AF000513; MID:g6459692; FIDN:AALF11467.1; PID:g64548
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1913
A:Map position: 1
C:Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain A; phage T4 DNA topoisomerases

Query Match 12.1%; Score 68.5; DB 2; Length 812;
Best Local Similarity 27.0%; Pred. No. 29;
Matches 27; Conservative 21; Mismatches 39; Indels 13; Gaps 5;

Oy 1 MSLYPATNIYITPLNLKGTTI-VNNVGVMKFFK-PPYLSKGTDCYSWT--IVDOTNRK 56
|:||| | : : : | : ||| :|| : |||
Db 599 MALVPG-----NDDEGELLAIISCEGLKRKVSDYPSKRGCLGVITLDVTRTYGKL 750

Oy 57 LTLCFLSSGEALPIITYKNGDIVRFHRLKIQQVYKKETOGI 96
| :|| :|| : :|| :|| :|| :||
Db 751 VTIAHVAGN-EELMLVTEKGIIVIRTVEEVRVTVGRNAQGV 789

RESULT 14
B83723
hypothetical protein BH0596 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C:Accession: B83723
R:Bakani, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, N.; Masui, N.; Fujii, F.; Hagiwara, S.
Nucleic Acids Res. 28: 4317-4331, 2000
A:title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans C-125
A:Reference number: AB3650; MUID:20263314
A:Accession: B83723
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-506 <STO>
A:Cross-references: GB:BA001509; GB:BA0000004; MID:g10173176; FIDN:BAR04305.1; GSFDRC:G10173176
A:Experimental source: strain C-125
C:Genetics:
C:Gene: BH0596

Query Match 11.9%; Score 67.5; DB 2; Length 306;
Best Local Similarity 29.2%; Pred. No. 11;
Matches 25; Conservative 14; Mismatches 28; Indels 21; Gaps 5;

Oy 8 NYIYIPLNOLKGGTJ-----VWVYGVVKKFKPPY-----LSKGTIDYCSVVIIVDDOTNKK 55
Db 219 NYMHL-LDHLKANTIDAIVNIDEIGVTPFNVOPLTSEIALSYETETEA VVVVIREDNKK 277
Oy 57 LTCLLFSGNYEALPII-----YKNGDIV 79
Db 278 LDVIL---NLISLPTISEVOAKVENDLI 303

RESULT 15

T21355
hypothetical protein F25H5.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21355
P:Steward, C.
submitted to the EMBL Data Library, October 1995
A:Reference number: Z19412
A:Accession: T21355
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-490 <WIL>
A:Cross-references: EMBL:ZP1068; PIDN:CAR02988.1; GSPDB:GN00019; CESP:F25H5.7
A:Experimental source: clone F25H5
C:Genetics:
A:Gene: CESP:F25H5.7
A:Map position: 1
A:Introns: 32/3; 174/3; 390/1; 440/3; 473/2

Query Match 11.8%; Score 67; DB 2; Length 490;
Best Local Similarity 23.4%; Pred. No. 23;
Matches 25; Conservative 23; Mismatches 47; Indels 12; Gaps 4;

Oy 8 NYIYIPLNOLKG---GTIVNVYGVVKKFKPPYLSKGTIDYCSVVIIVDDOINVKLTCLLFSG 64
Db 228 DVIHASWRMPGCKDQSTVITTOGPL---PEILS---DFWHMIYOEKIAVYVLMCLTFEG 280
Oy 65 NYEALPIIY--KNGDIVRHLKIOVKKEITOGITSSGFASLIPEGI 109
Db 281 GVEKCALYPERKLGWVRFGKYEITLTCKEEAIAGIWNLIVTINT 327

Search completed: April 12, 2002, 08:33:59
Job time: 43 sec

12	69.5	12.3	1161	21	AA51666	A. aeolicus AASEQ5
13	67	11.8	1256	13	AA27746	Muramidase release
14	66	11.6	336	19	AA98449	H. pylori GPO 578
15	66	11.6	1254	11	AA07503	Merzite apical-en
16	66	11.6	1254	18	AA24575	Merozoite apical-e
17	65.5	11.4	579	21	AA55702	Cosmid cHRM5 enco
18	65	11.5	221	21	AA02003	Arabidopsis thalia
19	65	11.5	221	21	AA02003	Arabidopsis thalia
20	65	11.5	221	21	AA02003	Arabidopsis thalia
21	65	11.5	221	21	AA02003	Arabidopsis thalia
22	65	11.5	221	21	AA02003	Arabidopsis thalia
23	65	11.5	221	21	AA02003	Arabidopsis thalia
24	65	11.5	221	21	AA02003	Arabidopsis thalia
25	65	11.5	221	21	AA02003	Arabidopsis thalia
26	65	11.5	221	21	AA02003	Arabidopsis thalia
27	65	11.5	221	21	AA02003	Arabidopsis thalia
28	65	11.5	221	21	AA02003	Arabidopsis thalia
29	65	11.5	221	21	AA02003	Arabidopsis thalia
30	65	11.5	221	21	AA02003	Arabidopsis thalia
31	65	11.5	221	21	AA02003	Arabidopsis thalia
32	65	11.5	221	21	AA02003	Arabidopsis thalia
33	65	11.5	221	21	AA02003	Arabidopsis thalia
34	65	11.5	221	21	AA02003	Arabidopsis thalia
35	65	11.5	221	21	AA02003	Arabidopsis thalia
36	65	11.5	221	21	AA02003	Arabidopsis thalia
37	65	11.5	221	21	AA02003	Arabidopsis thalia
38	65	11.5	221	21	AA02003	Arabidopsis thalia
39	65	11.5	221	21	AA02003	Arabidopsis thalia
40	65	11.5	221	21	AA02003	Arabidopsis thalia
41	65	11.5	221	21	AA02003	Arabidopsis thalia
42	65	11.5	221	21	AA02003	Arabidopsis thalia
43	65	11.5	221	21	AA02003	Arabidopsis thalia
44	65	11.5	221	21	AA02003	Arabidopsis thalia
45	65	11.5	221	21	AA02003	Arabidopsis thalia

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.
OM protein - protein search, using sw model
Run on: April 12, 2002, 08:33:16 ; Search time 23.56 seconds
(without alignments)
341.251 Million cell updates/sec
Title: US-09-816-248-5
Perfect score: 567
Sequence: 1 MSYPAIYVYPLNQLKGG.....KKEIQGIFSSGFASLIFLFI 109
Scoring table: BLOSUM62
Gap 10.0 , Gapext 0.5
Searched: 522463 seqs. 74073290 residues
Total number of hits satisfying chosen parameters: 522463
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : A_Geneseq_1101.*
1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAI.*
2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAI.*
3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAI.*
4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAI.*
5: /SID58/gcgdata/geneseq/geneseq/AA1984.DAI.*
6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAI.*
7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAI.*
8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAI.*
9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAI.*
10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAI.*
11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAI.*
12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAI.*
13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAI.*
14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAI.*
15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAI.*
16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAI.*
17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAI.*
18: /SID58/gcgdata/geneseq/geneseq/AA1997.DAI.*
19: /SID58/gcgdata/geneseq/geneseq/AA1998.DAI.*
20: /SID58/gcgdata/geneseq/geneseq/AA1999.DAI.*
21: /SID58/gcgdata/geneseq/geneseq/AA2000.DAI.*
22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAI.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

us-09-816-248-5.rag

XX	26-JUL-2001.	2000US-0488725.	XX
XX	25-APR-2000;	2000US-052317.	XX
XX	25-APR-2000;	2000US-0598042.	XX
XX	19-JUL-2000;	2000US-0620312.	XX
XX	19-JUL-2000;	2000US-0653450.	XX
XX	14-SEP-2000;	2000US-0662191.	XX
XX	19-OCT-2000;	2000US-0693035.	XX
XX	29-NOV-2000;	2000US-0727344.	XX
XX	26-DEC-2000;	2000WO-US34253.	XX
XX	21-JAN-2000;	2000US-0488725.	XX
XX	25-APR-2000;	2000US-052317.	XX
XX	03-JUL-2000;	2000US-0598042.	XX
XX	19-JUL-2000;	2000US-0620312.	XX
XX	03-AUG-2000;	2000US-0653450.	XX
XX	14-SEP-2000;	2000US-0662191.	XX
XX	19-OCT-2000;	2000US-0693035.	XX
XX	29-NOV-2000;	2000US-0727344.	XX

RESULT	1						
ID	AA40125	AA40125	standard: Protein: 634 AA.				
XX	AA40125:						
XX	22-OCT-2001	{first entry}					
XX	Human polypeptide	SEQ ID NO 3270.					
XX	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;						
XX	peripheral nervous system; neuropathy; central nervous system; CNS;						
XX	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;						
XX	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;						
XX	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;						
XX	leukaemia.						
XX	Homo sapiens.						
XX	W0200153312-A1.						
XX	26-JUL-2001.						
XX	26-DEC-2000:	2000WO-US34263.					
XX	21-JAN-2000:	2000US-0488725.					
XX	25-APR-2000:	2000US-0552317.					
XX	03-JUL-2000:	2000US-0598042.					
XX	19-JUL-2000:	2000US-0620312.					
XX	03-AUG-2000:	2000US-0653450.					
XX	14-SEP-2000:	2000US-0682191.					
XX	19-OCT-2000:	2000US-0693036.					
XX	29-NOV-2000:	2000US-0727344.					

ALIGNMENTS

```

PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yan Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac PI;
XX
DR WPI: 2001-442253/47.
XX
DR N-PSDB: AA159281.
XX
PI Novel nucleic acids and polypeptides, useful for treating disorders
PI and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
PS Example 5: SEQ ID NO 3270; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AA138642-AA142213) with neurotropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 534 AA:

Query Match 100.0%; Score 567; DB 22; Length 534;
Best Local Similarity 100.0%; Pred. No. 5.6e-63;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0:

QY 1 MSLVPATNYIYTPNLKGGTIVNVYGVVKFKPPYLSKGTDCSVVTIVDQTNVKLTCL 50
DB 1 mslvpacnyitpnlqkgtcivnvvgvkvfkppylskgtcysvvtivdqcnvklctcl 50
QY 61 LFSGNYEALPIIYKNGDIVRFRHLKTOVYKKETQGITSSGFASLTFEGT 109
DB 61 lfsqnyealpiiykngdivrfrhlkivvykktqgitssgfaltfegt 109

RESULT 2
AAB92742
ID AAB92742 standard; Protein: 534 AA.
AC AAB92742:
XX
XX 26-JUN-2001 (first entry)
XX
XX Human protein sequence SEQ ID NO:11188.
XX
XX Human: primer: detection; diagnosis; antisense therapy; gene therapy.
XX
XX Homo sapiens.
XX
XX EP1074517-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 92JP-0248036.
XX
XX 27-AUG-1999; 99JP-0300253.
XX
XX 11-JAN-2000; 2000JP-0118776.
XX
XX 02-MAY-2000; 2000JP-0183767.
XX
XX 09-JUN-2000; 2000JP-0241859.
XX

```

```

PA (HELL-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI: 2001-318749/34.
XX
XX
PI Primer sets for synthesizing polynucleotides, particularly the 5602
PI full-length cDNAs defined in the specification, and for the detection
PI and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
PS Claim 9: SEQ ID 11188; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
XX Sequence 534 AA:

Query Match 100.0%; Score 567; DB 22; Length 534;
Best Local Similarity 100.0%; Pred. No. 5.6e-63;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0:

QY 1 MSLVPATNYIYTPNLKGGTIVNVYGVVKFKPPYLSKGTDCSVVTIVDQTNVKLTCL 50
DB 1 mslvpacnyitpnlqkgtcivnvvgvkvfkppylskgtcysvvtivdqcnvklctcl 50
QY 61 LFSGNYEALPIIYKNGDIVRFRHLKTOVYKKETQGITSSGFASLTFEGT 109
DB 61 lfsqnyealpiiykngdivrfrhlkivvykktqgitssgfaltfegt 109

RESULT 3
AAB93478
ID AAB93478 standard; Protein: 534 AA.
XX
XX AAB93478:
XX
XX 26-JUN-2001 (first entry)
XX
XX Human protein sequence SEQ ID NO:12761.
XX
XX Human: primer: detection; diagnosis; antisense therapy; gene therapy.
XX
XX Homo sapiens.
XX
XX EP1074517-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX

```

```

FR 29-JUL-1999: 99JP-0248036.
PR 27-AUG-1999: 99JP-0300253.
PR 11-JAN-2000: 2000JP-0118776.
PR 02-MAY-2000: 2000JP-0183767.
PR 09-JUN-2000: 2000JP-0241899.
XX
PR (HELI-) HELIX RES INST.
PA
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI: 2001-318749/34.
XX
PI Primer sets for synthesizing polynucleotides, particularly the 5602
PI full-length cDNAs defined in the specification, and for the detection
PI and/or diagnosis of the abnormality of the proteins encoded by the
PI full-length cDNAs.
XX
PS Claim 8: SEQ ID 12761: 2537pp + CD ROM: English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primers sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92445 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 534 AA:
;

Query Match 100.0%; Score 567; DB 22: Length 634;
Best Local Similarity 100.0%; Pred. No. 5.6e-63;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSLVPATNYIYPLNQLKGGTIWNVYGVVKKFPPYLSKGTDCSVVTVVDOTNKLITCL 60
DB 1 mslvpatnyiytpnlqkkggtiwnvygvvkkfppylskgtid;csvvti;vdotnklitcl 60

OY 61 LFSGNYEALPIIYKNGDIVRFHRLKIQVYKKEQTGGITSSGFSALTFEET 109
DB 61 lfsenyaealpiiykngdivrfhrlikqvyykktggtitssgsfalsltfeet 109

RESULT 4
AAB95120
ID AAB95120 standard; Protein: 534 AA.
XX
AC AAB95120;
XX
DI 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:17113.
XX
EW Human: primer: detection; diagnosis: antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX

```

```

FN EPI074517-A2.
XX
FD 07-FEB-2001.
XX
PF 28-JUL-2000: 2000EP-0116126.
XX
PR 22-JUL-1999: 99JP-0248036.
PR 27-AUG-1999: 99JP-0300253.
PR 11-JAN-2000: 2000JP-0118776.
PR 02-MAY-2000: 2000JP-0183767.
PR 09-JUN-2000: 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI: 2001-318749/34.
XX
PI Primer sets for synthesizing polynucleotides, particularly the 5602
PI full-length cDNAs defined in the specification, and for the detection
PI and/or diagnosis of the abnormality of the proteins encoded by the
PI full-length cDNAs.
XX
PS Claim 8: SEQ ID 17113: 2537pp + CD ROM: English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primers sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 534 AA:

```

DE Arabidopsis thaliana protein fragment SEQ ID NO: 75533.
XX Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX EP1033405-A2.
PN
PD 06-SEP-2000.
XX
PF 25-FEB-2000: 2000EP-0301439.
XX
PR 25-FEB-1999: 99US-0121825.
PR 05-MAR-1999: 99US-0123180.
PR 09-MAR-1999: 99US-0123548.
PR 23-MAR-1999: 99US-0125788.
PR 25-MAR-1999: 99US-0125254.
PR 29-MAR-1999: 99US-0125785.
PR 01-APR-1999: 99US-0127462.
PR 05-APR-1999: 99US-0128234.
PR 08-APR-1999: 99US-0128714.
PR 16-APR-1999: 99US-0129845.
PR 19-APR-1999: 99US-0130077.
PR 21-APR-1999: 99US-0130449.
PR 23-APR-1999: 99US-0130510.
PR 28-APR-1999: 99US-0130891.
PR 30-APR-1999: 99US-0131449.
PR 30-APR-1999: 99US-0132048.
PR 30-APR-1999: 99US-0132407.
PR 04-MAY-1999: 99US-0132484.
PR 05-MAY-1999: 99US-0132485.
PR 06-MAY-1999: 99US-0132486.
PR 06-MAY-1999: 99US-0132487.
PR 07-MAY-1999: 99US-0132863.
PR 11-MAY-1999: 99US-0133456.
PR 14-MAY-1999: 99US-0134218.
PR 14-MAY-1999: 99US-0134219.
PR 14-MAY-1999: 99US-0134221.
PR 14-MAY-1999: 99US-0134370.
PR 18-MAY-1999: 99US-0134768.
PR 19-MAY-1999: 99US-0134941.
PR 20-MAY-1999: 99US-0135124.
PR 21-MAY-1999: 99US-0135353.
PR 24-MAY-1999: 99US-0135629.
PR 25-MAY-1999: 99US-0136021.
PR 27-MAY-1999: 99US-0136392.
PR 28-MAY-1999: 99US-0136782.
PR 01-JUN-1999: 99US-0137222.
PR 03-JUN-1999: 99US-0137528.
PR 04-JUN-1999: 99US-0137502.
PR 07-JUN-1999: 99US-0137724.
PR 08-JUN-1999: 99US-0138094.
PR 10-JUN-1999: 99US-0138540.
PR 10-JUN-1999: 99US-0138847.
PR 14-JUN-1999: 99US-0139119.
PR 16-JUN-1999: 99US-0139452.
PR 16-JUN-1999: 99US-0139453.
PR 17-JUN-1999: 99US-0139492.
PR 18-JUN-1999: 99US-0139454.
PR 18-JUN-1999: 99US-0139455.
PR 18-JUN-1999: 99US-0139456.
PR 18-JUN-1999: 99US-0139457.
PR 18-JUN-1999: 99US-0139458.
PR 18-JUN-1999: 99US-0139459.
PR 18-JUN-1999: 99US-0139460.
PR 18-JUN-1999: 99US-0139461.
PR 18-JUN-1999: 99US-0139462.
PR 18-JUN-1999: 99US-0139463.
PR 18-JUN-1999: 99US-0139464.
PR 18-JUN-1999: 99US-0139465.
PR 18-JUN-1999: 99US-0139466.
PR 18-JUN-1999: 99US-0139467.
PR 18-JUN-1999: 99US-0139468.
PR 18-JUN-1999: 99US-0139469.
PR 18-JUN-1999: 99US-0139470.
PR 21-JUN-1999: 99US-0139817.
PR 22-JUN-1999: 99US-0139899.
PR 23-JUN-1999: 99US-0140353.
PR 23-JUN-1999: 99US-0140354.
PR 24-JUN-1999: 99US-0140695.
PR 28-JUN-1999: 99US-0140823.
PR 29-JUN-1999: 99US-0140991.
PR 30-JUN-1999: 99US-0141287.
PR 01-JUL-1999: 99US-0141842.
PR 01-JUL-1999: 99US-0142154.
PR 02-JUL-1999: 99US-0142055.
PR 06-JUL-1999: 99US-0142390.
PR 08-JUL-1999: 99US-0142803.
PR 09-JUL-1999: 99US-0142920.
PR 12-JUL-1999: 99US-0142977.
PR 13-JUL-1999: 99US-0143542.
PR 14-JUL-1999: 99US-0143624.
PR 15-JUL-1999: 99US-0144005.
PR 16-JUL-1999: 99US-0144085.
PR 16-JUL-1999: 99US-0144086.
PR 19-JUL-1999: 99US-0144325.
PR 19-JUL-1999: 99US-0144331.
PR 19-JUL-1999: 99US-0144332.
PR 19-JUL-1999: 99US-0144333.
PR 19-JUL-1999: 99US-0144334.
PR 19-JUL-1999: 99US-0144335.
PR 20-JUL-1999: 99US-0144352.
PR 20-JUL-1999: 99US-0144632.
PR 20-JUL-1999: 99US-0144884.
PR 21-JUL-1999: 99US-0144814.
PR 21-JUL-1999: 99US-0145086.
PR 21-JUL-1999: 99US-0145088.
PR 22-JUL-1999: 99US-0145085.
PR 22-JUL-1999: 99US-0145087.
PR 22-JUL-1999: 99US-0145089.
PR 22-JUL-1999: 99US-0145192.
PR 23-JUL-1999: 99US-0145145.
PR 23-JUL-1999: 99US-0145218.
PR 23-JUL-1999: 99US-0145224.
PR 26-JUL-1999: 99US-0145276.
PR 27-JUL-1999: 99US-0145913.
PR 27-JUL-1999: 99US-0145918.
PR 27-JUL-1999: 99US-0145919.
PR 28-JUL-1999: 99US-0145951.
PR 02-AUG-1999: 99US-0146386.
PR 02-AUG-1999: 99US-0146388.
PR 02-AUG-1999: 99US-0146389.
PR 03-AUG-1999: 99US-0147038.
PR 04-AUG-1999: 99US-0147204.
PR 04-AUG-1999: 99US-0147302.
PR 05-AUG-1999: 99US-0147192.
PR 05-AUG-1999: 99US-0147260.
PR 05-AUG-1999: 99US-0147303.
PR 05-AUG-1999: 99US-0147416.
PR 09-AUG-1999: 99US-0147493.
PR 09-AUG-1999: 99US-0147935.
PR 10-AUG-1999: 99US-0148171.
PR 11-AUG-1999: 99US-0148319.
PR 12-AUG-1999: 99US-0148341.
PR 13-AUG-1999: 99US-0148565.
PR 13-AUG-1999: 99US-0148684.
PR 16-AUG-1999: 99US-0149368.
PR 17-AUG-1999: 99US-0149175.
PR 18-AUG-1999: 99US-0149426.
PR 20-AUG-1999: 99US-0149722.
PR 20-AUG-1999: 99US-0149723.
PR 20-AUG-1999: 99US-0149929.
PR 23-AUG-1999: 99US-0149902.
PR 23-AUG-1999: 99US-0149930.
PR 25-AUG-1999: 99US-0150566.
PR 25-AUG-1999: 99US-0150884.
PR 26-AUG-1999: 99US-0151065.
PR 27-AUG-1999: 99US-0151066.
PR 27-AUG-1999: 99US-0151080.


```

PR 30-AUG-1999; 99US-0151303;
PR 31-AUG-1999; 99US-0151438;
PR 01-SEP-1999; 99US-0151930;
PR 07-SEP-1999; 99US-0152363;
PR 10-SEP-1999; 99US-0153070;
PR 13-SEP-1999; 99US-0153758;
PR 15-SEP-1999; 99US-0154018;
PR 16-SEP-1999; 99US-0154039;
PR 20-SEP-1999; 99US-0154779;
PR 22-SEP-1999; 99US-0155139;
PR 23-SEP-1999; 99US-0155486;
PR 24-SEP-1999; 99US-0155659;
PR 28-SEP-1999; 99US-0156458;
PR 29-SEP-1999; 99US-0156595;
PR 04-OCT-1999; 99US-0157117;
PR 05-OCT-1999; 99US-0157753;
PR 06-OCT-1999; 99US-0157855;
PR 07-OCT-1999; 99US-0158029;
PR 08-OCT-1999; 99US-0158232;
PR 12-OCT-1999; 99US-0158369;
PR 13-OCT-1999; 99US-0158293;
PR 13-OCT-1999; 99US-0158294;
PR 13-OCT-1999; 99US-0158295;
PR 14-OCT-1999; 99US-0159329;
PR 14-OCT-1999; 99US-0159330;
PR 14-OCT-1999; 99US-0159331;
PR 14-OCT-1999; 99US-0159637;
PR 14-OCT-1999; 99US-0159638;
PR 18-OCT-1999; 99US-0159584;
PR 21-OCT-1999; 99US-0160741;
PR 21-OCT-1999; 99US-0160767;
PR 21-OCT-1999; 99US-0160768;
PR 21-OCT-1999; 99US-0160770;
PR 21-OCT-1999; 99US-0160814;
PR 21-OCT-1999; 99US-0160815;
PR 22-OCT-1999; 99US-0160980;
PR 22-OCT-1999; 99US-0160981;
PR 22-OCT-1999; 99US-0160989;
PR 25-OCT-1999; 99US-0161404;
PR 25-OCT-1999; 99US-0161405;
PR 25-OCT-1999; 99US-0161406;
PR 25-OCT-1999; 99US-0161359;
PR 26-OCT-1999; 99US-0161360;
PR 26-OCT-1999; 99US-0161361;
PR 28-OCT-1999; 99US-0161920;
PR 28-OCT-1999; 99US-0161922;
PR 28-OCT-1999; 99US-0161993;
PR 29-OCT-1999; 99US-0162142;

Query Match 15.4%; Score 87.5; DB 21; Length 487;
Best Local Similarity 31.2%; Pred. No. 0.019;
Matches 29; Conservative 16; Mismatches 39; Indels 9; Gaps 4;

QY 23 VNYGVVYKFFPPVLSKGDYCSVTVDDT--NVKLTCLLFGSGNVEALPIIYKNDIVR 90
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 24 vsliqlvleqrepkqrndwictlrlidtdtyspgltv:nv:fskltelprv:shddmil 83

QY 81 FHLKIOYK--KETOGITS--SGFASLTFEG 108
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 84 ftlikmqfdsgkrvnaa:rw:ssfa--lfeg 114

```

RESULT 6
 ID AAY13942
 AC AAY13942 standard; Protein: 383 AA.

XX AAY13942;

XX 14-JUL-1999 (first entry)

XX Human transmembrane protein, HP01737.

XX Transmembrane protein; human; cell membrane; proliferation; diagnosis;

KW cell differentiation; carcinostatic agent; probe; gene therapy;
 KW signal transduction; apoptosis; inhibitor;
 KW phosphatidylethanolamine N-methyltransferase.

XX Homo sapiens.

PN W09918203-A2.

PD 15-APR-1999.

PF 05-OCT-1998; 98WO-JP04475.

PP 08-OCT-1997; 97JP-0275271.

XX (PROT-) PROTEGENE INC.

PA (SAGA) SAGAMI CHEM RES CENT.

XX Kato S, Kobayashi M, Sekine S, Yamaguchi T;

XX KPI: 1999-277268/23.

DR N-ESDB: AAX36808, AAX36809.

XX Human transmembrane proteins and nucleotide sequences

PS Claim 1: Page 87-89; 139pp; English.

XX This sequence is a human transmembrane protein of the invention.
 CC All of the proteins exist in the cell membrane, so are considered to be
 CC proteins controlling the proliferation and differentiation of the cells.
 CC They may be useful as carcinostatic agents or as antigens for preparing
 CC antibodies against the proteins. The cDNAs can be used as probes for
 CC gene diagnosis and gene sources for gene therapy, as well as for
 CC large-scale expression of the proteins. The HP01498 (see AAY13939)
 CC protein may be associated with signal transduction associated with
 CC apoptosis, and therefore useful in inhibition of apoptosis. The HP01952
 CC (see AAY13943) protein can be used to treat diseases associated with
 CC phosphatidylethanolamine N-methyltransferase. The proteins are
 CC identified by the presence of a hydrophobic transmembrane region,
 CC knowledge of the protein function is not required, as in e.g. methods of
 CC expression cloning.

SO Sequence 383 AA;

Query Match 13.1%; Score 74; DB 20; Length 383;

Best Local Similarity 26.2%; Pred. No. 0.75;

Matches 17; Conservative 18; Mismatches 26; Indels 4; Gaps 2;

QY 36 YLSKGTDCYSVTVDDTQNVKL--TCLLFGSGNVEALPIIYK--NGDIVFHLKIOYK 91

Db 248 hlsfgedypgilvnpldhntvtapqasmmtqfykvvptvymkvdgvlrtngfsvtrhek 307

QY 92 ETGGI 96

Db 308 vangel 312

RESULT

AAG75511

ID AAG75511 standard; Protein: 410 AA.

XX AAG75511;

XX Q3-SEP-2001 (first entry)

XX Human colon cancer antigen protein SEQ ID NO:6375.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;

XX colorectal carcinoma; chromosome 2.

XX Homo sapiens.

XX W0200122320-A2.

XX PD 05-APR-2001.
XX PF 28-SEP-2000: 2000WO-US26524.
XX PR 29-SEP-1999: 99US-0157137.
XX PR 03-NOV-1999: 99US-0163280.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Ruben SM, Barash SC, Birse CE, Rosen CA:
XX PI WPI: 2001-235357/24.
XX DR N-PSDB: AAH35016.
XX XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX PS Claim 11: Page 7845-7847: 9803pp: English.
XX XX AAH32943 to AAH37195 and AAH37788 represent human colon
XX CC cancer-associated nucleic acid molecules (N) and proteins (P), where
XX CC the proteins are collectively known as colon cancer antigens. The colon
XX CC cancer antigens have cytostatic activity and can be used in gene
XX CC therapy and vaccine production. N and P may be used in the prevention,
XX CC diagnosis and treatment of diseases associated with inappropriate P
XX CC expression. For example, N and P may be used to treat disorders
XX CC associated with decreased expression by rectifying mutations or deletions
XX CC in a patient's genome that affect the activity of P by expressing
XX CC inactive proteins or to supplement the patient's own production of P.
XX CC Additionally, N may be used to produce the colon cancer-associated Ps
XX CC by inserting the nucleic acids into a host cell and culturing the cell
XX CC to express the proteins. N and P can be used in the prevention, diagnosis
XX CC and treatment of colorectal carcinomas and cancers. AAH37195 to AAH37204
XX CC and AAH37789 represent sequences used in the exemplification of the
XX CC present invention.
XX CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX CC missing at time of publication, meaning no sequences are present for
XX CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX XX Sequence 410 AA:
SO
Query Match 13.1% Score 74; DB 22; Length 410;
Best Local Similarity 26.2%; Pred. No. 0.82; Mismatches 4; Gaps 2;
Matches 17; Conservative 18; Indels 4; Gaps 2;
OY 36 YLSKGDYCSVVIIVDTQINVKL--TCLFSGNVEALPIIYK--NGDIVRPHRLKIOVYK 91
DB 275 hlsfgedypgviupldhtn:tapqasmfgyfkvypvymk:dge:lrtngqsvt:hek 334
OY 92 ETQGI 96
DB 335 vangl 339
RESULT 8
AAW51088 standard: Protein: 287 AA.
XX AC AAW51088;
XX XX 14-SEP-1998 (first entry)
XX DE Cowdria ruminantium major antigenic protein 1 (MAP1).
XX KW MAP1 gene; major antigenic protein 1; rickettsia; heartworm;
XX KW DNA vaccine.
XX XX Cowdria ruminantium.
XX OS WC2816554-A1.
XX PN (UYFL) UNIV FLORIDA.
XX XX

PD 23-APR-1998.
XX PF 17-OCT-1997: 97WO-US19044.
XX PR 17-OCT-1996: 96US-0733230.
XX PA (UYFL) UNIV FLORIDA.
XX PI Barbet AF, Burridge MJ, Canta RR, Mahan SM, McGuire TC:
XX PI Nyika A, Rurangirwa FR:
XX DR WPI: 1998-251232/22.
XX DR N-PSDB: AAV07176.
XX XX Composition containing nucleic acid encoding rickettsial antigen -
XX PT useful for, e.g. stimulating protective immune response in humans or
XX PT animals
XX PS Claim 3: Page 15-16: 39pp: English.
XX XX This polypeptide comprises the major antigen protein 1 gene (MAP1)
XX CC of Cowdria ruminantium, the causative agent of heartwater in
XX CC domestic ruminants. It is encoded by the MAP1 gene (see AAV07176).
XX CC A claimed composition comprises a nucleic acid (see AAV07176-82)
XX CC encoding a polypeptide (see AAW51088-99) that elicits a protective
XX CC immune response against a rickettsial pathogen. The nucleic acid
XX CC is used, in human or veterinary medicine, in vaccines to protect
XX CC against Rickettsia, Ehrlichia, Anaplasma and Cowdria species. The
XX CC nucleic acid does not replicate in the host but remains episomal
XX CC and capable of expressing polypeptide for at least 19 mth.
XX XX Sequence 287 AA:
SO
Query Match 12.4% Score 70.5; DB 19; Length 287;
Best Local Similarity 30.3%; Pred. No. 1.4;
Matches 27; Conservative 19; Mismatches 18; Indels 25; Gaps 7;
OY 35 PYLSK--GTDYCSVVIIVDTQINVKLTC-----LLFSGNVEALPIIYKNGDIVRPHRL--- 84
DB 194 pyvcagiqtd---lvsvinatnpklsyqgklqisynsea--sifigg---hfrhrgn 245
OY 85 -----KIOVYKKEQGITSSGFASLIFE 107
DB 246 efkdlatikfiskt-gisnpqfasatid 273
RESULT 9
AAB36182 standard: Protein: 287 AA.
XX AC AAB36182;
XX XX 02-MAR-2001 (first entry)
XX DE Cowdria ruminantium MAP1.
XX XX Cowdria ruminantium: MAP1: major antigenic protein 1; antirickettsial;
XX KW vaccine; gene therapy; Rickettsia; Ehrlichia; Anaplasma; map2: lhworf3;
XX KW 4hworfl: 18hworfl: 3gdorf3.
XX XX Cowdria ruminantium.
XX OS WC200005063-A2.
XX PN 02-NOV-2000.
XX PD 21-APR-2000: 2000WO-US10886.
XX PE 22-APR-1999: 99US-0130725.
XX PF (UYFL) UNIV FLORIDA.
XX PA
XX XX

CC template-dependent elongation of nucleic acids which comprises a
 CC thermostable sliding clamp protein, which is connected with a
 CC elongation protein that shows thermostable polymerase activity. The
 CC thermostable in vitro accessory complex can be used to produce the
 CC thermostable in vitro complex, which is useful for template-dependent
 CC elongation of nucleic acids, e.g. for amplification or reverse
 CC transcription. This is useful for sequencing nucleic acids by the
 CC polymerase chain reaction or reverse transcriptase-polymerase chain
 CC reaction (RT-PCR). The complex can be used to mark nucleic acids.
 CC AAY52000-Y52084 and AAY90752-W90799 represent proteins and protein
 CC fragments used to illustrate the method of the invention.
 XX
 SQ Sequence 1161 AA:
 Query Match 12.3% Score 69.5; DB 21; Length 1161;
 Best Local Similarity 29.3%; Pred. No. 13;
 Matches 22; Conservative 12; Mismatches 24; Indels 17; Gaps 4:
 OY 5 PATNY-----IVTLPNOLK-----GGTIVNVYGVVKKFPPLSKGTIVSVVLTVD 52
 DB 972 pldnyekllkrrytpiedleekeseavltgvtelkxk-----tkngdy-mavfnlvdk 1027
 OY 53 INVKLTCLLFSGNYE 67
 DB 1028 tgl-icrrvifpgrye 1041
 RFSULT 12
 ID AAY51566 standard; protein; 1161 AA.
 XX AAY51566:
 AC AAY51565:
 XX 01-JUN-2000 (first entry)
 DE A. aeolicus AAY5050 protein fragment.
 KW Thermotable; template-dependent elongation; staple protein;
 KW elongation protein; amplification; reverse transcription.
 XX Aquifex aeolicus.
 XX W0200008164-A2.
 XX 17-FEB-2000.
 XX 06-AUG-1999; 99WO-DF02480.
 XX 06-AUG-1998; 99DE-1035653.
 XX 07-SEP-1998; 98DE-1040771.
 XX 18-JUN-1993; 99EP-0111795.
 XX (LION-) LION BIOSCIENCE AG.
 XX Kilger C., Kober I., Voss H., Moeckel G;
 XX WPI: 2000-195576/17.
 XX Thermotable in vitro polymerase complex for template-dependent
 XX elongation of nucleic acids
 XX Disclosure: Page 223-227; 233pp; German.
 XX This invention describes a novel thermostable in vitro complex (A) for
 CC template-dependent elongation of nucleic acids (NA) comprising a
 CC thermostable "staple" protein and an elongation protein. The thermostable
 CC in vitro complex is useful for template-dependent elongation of NA, e.g.
 CC for amplification or reverse transcription. This is useful for sequencing
 CC NA by the polymerase chain reaction or reverse transcriptase-PCR. The
 CC complex can be used to mark NA. This sequence represents the Aquifex
 CC aeolicus AAY5050 protein which is used to illustrate the method of the
 CC invention.

XX SQ Sequence 1161 AA:
 Query Match 12.3% Score 69.5; DB 21; Length 1161;
 Best Local Similarity 29.3%; Pred. No. 13;
 Matches 22; Conservative 12; Mismatches 24; Indels 17; Gaps 4:
 OY 5 PATNY-----IVTLPNOLK-----GGTIVNVYGVVKKFPPLSKGTIVSVVLTVD 52
 DB 972 pldnyekllkrrytpiedleekeseavltgvtelkxk-----tkngdy-mavfnlvdk 1027
 OY 53 INVKLTCLLFSGNYE 67
 DB 1028 tgl-icrrvifpgrye 1041
 RFSULT 13
 ID AAR27746 standard; Protein: 1256 AA.
 XX AAR27746:
 AC AAR27745:
 XX 04-MAR-1993 (first entry)
 DE Muramidase released protein.
 KW MRP: detection; prevention; screening; diagnostic.
 XX Streptococcus suis type II pathogenic.
 XX Key Location/Qualifiers
 FI Peptide 1...47
 FI Peptide /note= "signal peptide"
 FI Peptide 48...1256
 FI Peptide /note= "mature peptide"
 FI Region 857...952
 FI Peptide /note= "proline-rich region"
 FI Peptide 1222...1256
 FI Peptide /note= "membrane anchor sequence"
 XX W02215630-A.
 XX 01-OCT-1992.
 XX 19-MAR-1992; 92WI-RL00054.
 XX 21-MAR-1991; 91HL-0000510.
 XX (DIPP-) CENT DIERGENESKUNDIG INST.
 XX Smith HE., Vecht U.
 XX WPI: 1992-349215/42.
 XX R-PSDB: AAQ29472.
 XX Deoxyribonucleic acid encoding virulence characteristic of
 XX Streptococcus suis - useful for antibody and polypeptide for
 XX diagnosing and preventing infections in pigs and humans
 XX Claim 9: Fig 2: 85pp; English.
 XX The sequence is that of the muramidase released protein from
 CC Streptococcus suis type II (pathogenic) which allows the detection
 CC and the prevention of infections by S. suis in a more effective
 CC manner than was previously possible. It facilitates screening of
 CC e.g. pigs and elimination of infected and carrier pigs can then be
 CC carried out. The new diagnostic tests can distinguish between
 CC avirulent and virulent strains. It may be used in the prodn. of
 CC a vaccine. See also AAR27744 and AAR27745.
 XX Sequence 1256 AA:
 SQ

```

Query Match          11.8%; Score 57; DB 13; Length 1255;
Best Local Similarity 23.2%; Pred. No. 29;
Matches 28; Conservative 15; Mismatches 33; Indels 20; Gaps 5;

Oy 7 INVIYPLNQLKGGIIVHWY-----GVVKFFKP----PYLSKGGIDYCSVVIIVDGIHWKL 57
Db 1000 tthyykktpeKnglvrvnyvtdgtvik--epvtdtptspegupj-----dt.dnkp 1050

Oy 58 ICLLFSG-NYEALPII-LKNGDIVRFHRLKIQVYKK 91
Db 1051 ktltfkgseyelrvdgtengkvvegetvvtvyryk 1086

RESULT 14
AAW98449
ID AAW98449 standard; Protein: 336 AA.
XX AC AAW98449:
XX DT 31-MAR-1999 (first entry)
XX DE H. pylori GHPO 578 protein.
XX DE GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
XX KW peptic ulcer disease.
XX KW Helicobacter pylori.
XX PN W09843478-AL.
XX PD 08-OCT-1998.
XX PF 01-APR-1998; 98SU-WU06371.
XX PR 29-JUL-1997; 97US-0902615.
XX PR 01-APR-1997; 97US-0833457.
XX PR 24-JUN-1997; 97US-0861227.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
XX DR WPI: 1998-542293/46.
XX DR N-PSDB: AAX14158.
XX
XX PT New isolated Helicobacter polynucleotides - used to develop products
XX PT for the diagnosis, prevention and treatment of Helicobacter
XX PT infections and gastrointestinal diseases
XX
XX PS Claim 8; Page 781-783; 2054pp; English.
XX
XX CC This sequence represents a Helicobacter pylori GHPO protein of the
XX CC invention. The polypeptides can be used for preventing or treating
XX CC Helicobacter infections, and gastroduodenal diseases associated with
XX CC these infections, including acute, chronic, and atrophic gastritis, and
XX CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
XX CC used for the production of antibodies. The products can also be used for
XX CC detection and diagnosis.
XX
XX SQ Sequence 336 AA;

```

```

Query Match      11.53; Score 66; DB 19; Length 336;
Best Local Similarity 35.44; Pred. NO. 6.5;
Matches 17; Conservative 8; Mismatches 17; Indels 10; Gaps 2;

OY 31 FFKPPLSKGDFCSVIIIDQINVKLTCLLFCSEYALPIYKNGDI 78
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 58 fggpgglak-tsihiakemetnikt-----aamlekkssdgl 95

```

[illegible]

PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs
XX
PS Claim 8: SEQ ID 11189; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 634 AA:

Query Match 100.0%; Score 3324; DB 22; Length 634;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 634; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSLVPAATNYITPLNQLKGGTIVNMGVWFFKPPYLSKGTDCYSVVTIVDQTNVLTCL 60
DB 1 mslvpaaatnyitplnqlkgtivnmgvwwffkppylskgtcysvvtivdqtntvltcl 60
QY 61 LPSGNYEALPIIYNGDITVRHRIKIOYKKETQGTSSGFSASITFPGTLAGAPLIPHTSS 120
DB 61 lpsgnyealpliiyngditvrhrikioykketqgtssgfsasitfpgttagapliptss 120
QY 121 KYENFTTDDHKMVAALRWASTHMSPSWTLLKLAIDVQPMQYFDTICOLLAKAEVDCASPL 180
DB 121 kyenfttddhkmvaelrwasthmspswtllkkladvqpmqyfdticollakaevdcaspl 180
QY 181 LKVMGDTRTPPSWRLIQDLVLEGLSHIHRLQNLITIDILVDNIIVIVARSLSKVGSLR 240
DB 181 lkvmgdtrtppswrliqdlvleglshihrlqnlitidilvdniivivarslskvgslr 240
QY 241 IYSLHTKLOSMMSENQTMISLFFHLHGCTSYGRCIRVLPESNSVDOLKKDLESANI,TAN 300
DB 241 iyslhtklosmmesenqtmislfhlhgctsygrcivr lpesnsvdolkkdlesanitan 300
QY 301 QHSDVLCQSEPDSPSSGVSLSYEVERCQQLSATITDHIQYLERTPICALIKKAKVQY 360
DB 301 qhsdvlcqsepdspssgvslysyeve rccqlsatitdhiqylertpicalikakvqy 360
QY 361 RIRAKLSRYKPRRLFOVKLHCPKCHLLOEVPHEGDLIDIFODGATKTPPVKLQNTSLYD 420
DB 361 riraklsrykprrlfovkhlhpkchllloevphegdlidifodgatktppvklqntsl 420
QY 421 SKTWTNKGKRVAVHVKNNGLPLNSNECLLLLEGGTSLSEICKLSKNKFSNIPVRSCH 480
DB 421 sktwtknkgkrvavhvknnlplnsneclllle ggtslseicklsknkfsnipvrsch 480
QY 481 EDLELLDLSAPFLIOCTIHGYCKQCSSLRSITONLSIVDKTSWTPSSVAALGLIVPQY 540
DB 481 edlelldlsapfliogctihgyckqcsslrsit onlsivdktswtpssvaalglivpqy 540
QY 541 VFVWTFLLDGGTGVLEAYLMDSDKFFQIPASEVLMDDDLQKSVDMIMDMFCPPGKIDAY 600

|||||
Db 541 vfmvtfllddgtgvleaylmdskdffqipasevlmdddlqksgvdmimdmfcppgikiday 600
QY 601 PWLECFIKSNVNTGTDNQICYOIFDTVAEDVI 634
|||||
Db 601 pwlecfiksnvntgtdnqicyoifdtctvaedvi 634

RESULT 2
AAM40125
ID AAM40125 standard; Protein: 634 AA.

XX AAM40125:

DT 22-OCT-2001 (first entry)

XX Human polypeptide SEQ ID NO 3270.

DE Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.

XX Homo sapiens.

OS WO200153312-A1.

PN 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-052317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0691036.

PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

PA Tian YT, Liu C, Asundi V, Chen R, Mu Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RC;

XX WPI: 2001-442253/47.

XX N-PSDB; AAI59281.

XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries.

XX Example 5: SEQ ID NO 3270; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using SW model

Run on: Apr 11 9, 2002, 16:58:21 ; Search time 58.75 seconds
(without alignments) 799.361 Million cell updates/sec

Title: US-09-816-248-13
Perfect score: 3324
Sequence: 1 MSLVATATVATYTPPLNOLKGC.....GTDNOLCVQIFPTVAHEVJ 634

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A.Geneseq_1101.*
2: /SIDSR/gcgdata/geneseq/geneseq/AA1980.DAT.*
3: /SIDSR/gcgdata/geneseq/geneseq/AA1981.DAT.*
4: /SIDSR/gcgdata/geneseq/geneseq/AA1982.DAT.*
5: /SIDSR/gcgdata/geneseq/geneseq/AA1983.DAT.*
6: /SIDSR/gcgdata/geneseq/geneseq/AA1984.DAT.*
7: /SIDSR/gcgdata/geneseq/geneseq/AA1985.DAT.*
8: /SIDSR/gcgdata/geneseq/geneseq/AA1986.DAT.*
9: /SIDSR/gcgdata/geneseq/geneseq/AA1987.DAT.*
10: /SIDSR/gcgdata/geneseq/geneseq/AA1988.DAT.*
11: /SIDSR/gcgdata/geneseq/geneseq/AA1989.DAT.*
12: /SIDSR/gcgdata/geneseq/geneseq/AA1990.DAT.*
13: /SIDSR/gcgdata/geneseq/geneseq/AA1991.DAT.*
14: /SIDSR/gcgdata/geneseq/geneseq/AA1992.DAT.*
15: /SIDSR/gcgdata/geneseq/geneseq/AA1993.DAT.*
16: /SIDSR/gcgdata/geneseq/geneseq/AA1994.DAT.*
17: /SIDSR/gcgdata/geneseq/geneseq/AA1995.DAT.*
18: /SIDSR/gcgdata/geneseq/geneseq/AA1996.DAT.*
19: /SIDSR/gcgdata/geneseq/geneseq/AA1997.DAT.*
20: /SIDSR/gcgdata/geneseq/geneseq/AA1998.DAT.*
21: /SIDSR/gcgdata/geneseq/geneseq/AA1999.DAT.*
22: /SIDSR/gcgdata/geneseq/geneseq/AA2000.DAT.*
23: /SIDSR/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3324	100.0	634	22	AAB92742
2	3317	99.8	634	22	AAB40125
3	3317	99.8	634	22	AAB3478
4	3317	99.8	634	22	AAB95120
5	1589	47.8	307	22	AA1911
6	358	10.8	66	22	AA193583
7	147	4.4	30	22	AA194264
8	138.5	4.2	467	21	AA19505
9	122	3.7	1144	17	AA198122
10	122	3.7	1144	17	AA198123
11	114	3.4	2391	15	AA195694

12	111	3.3	1139	21	AAB18793	Amino acid sequenc
13	106	3.2	1123	21	AAG35865	Arabidopsis thalia
14	105.5	3.2	1557	21	AAV84730	Amino acid sequenc
15	105.5	3.2	1574	21	AAV84729	Amino acid sequenc
16	105.5	3.2	1578	21	AAV84732	Amino acid sequenc
17	105.5	3.2	1588	21	AAV84731	Amino acid sequenc
18	105.5	3.2	1605	21	AAV84733	Amino acid sequenc
19	105	3.2	647	21	AAV57583	Human papillomavir
20	105	3.2	1112	21	AAV58113	Haemorrhagic enter
21	105	3.2	1516	21	AAV81819	Plasmodium falcipa
22	104	3.1	785	21	AAH03155	Human ubiquitin-sp
23	103.5	3.1	1002	21	AAH03155	Arabidopsis thalia
24	103.5	3.1	1266	18	AAW25157	Tomato immunity 2
25	103.5	3.1	1266	18	AAW25157	Amino acid sequenc
26	103.5	3.1	1568	20	AAV13462	Viral-encoded sema
27	103.5	3.1	1568	20	AAV13462	Viral-encoded sema
28	103.5	3.1	1568	21	AAV828522	Human VESPR. Homo
29	103.5	3.1	1568	22	AAV00239	Viral-encoded sema
30	103.5	3.1	1568	22	AAV00239	Human VESPR. Homo
31	103.5	3.1	1568	22	AAV0131	Viral-encoded sema
32	102.5	3.1	1568	22	AAV0131	Human VESPR. Homo
33	102.5	3.1	1568	22	AAV0131	Amino acid sequenc
34	101	3.0	611	22	AAH31889	Amino acid sequenc
35	101	3.0	611	22	AAH31889	S cerevisiae apopt
36	100.5	3.0	1604	19	AAW64179	Lettuce resistance
37	100.5	3.0	653	22	AAW64179	Human PP7. Homo s
38	100.5	3.0	1119	20	AAV08114	Human PKO326 prote
39	100.5	3.0	1119	21	AAV08114	Human PKO326 prote
40	100.5	3.0	1119	22	AAV08114	Human PKO326 poly
41	100.5	3.0	1119	22	AAV08114	Human PKO326 poly
42	100.5	3.0	1119	22	AAV08114	Human PKO326 poly
43	100.5	3.0	1119	22	AAV08114	Human PKO326 poly
44	100	3.0	1040	21	AAH48118	Amino acid sequenc
45	100	3.0	1059	21	AAH48117	Arabidopsis thalia

ALIGNMENTS

RESULT 1	
AAB92742	Standard; Protein; 634 AA.
ID	AAB92742
XX	26-JUN-2001 (first entry)
AC	Human protein sequence SEQ ID NO:11188.
XX	Human: primer: detection; diagnosis; antisense therapy; gene therapy.
XX	Homo sapiens.
OS	EP1074617-A2.
XX	07-FEB-2001.
XX	28-JUL-2000; 2000EP-0116126.
XX	29-JUL-1999; 99JP-0248036.
PR	27-AUG-1999; 99JP-0300253.
PR	11-JAN-2000; 2000JP-0118776.
PR	02-MAY-2000; 2000JP-0183767.
PR	09-JUN-2000; 2000JP-0241899.
PA	(HELI-) HELIX RES INST.
PI	De T. Isogai T. Nishikawa T. Hayashi K. Saito K. Yamamoto J.
PI	Ushii S. Sugiyama T. Wakamatsu A. Nagai K. Otsuki T.
DR	WPI: 2001-318749/34.
PR	Primer sets for synthesizing polynucleotides, particularly the 5602

```
Db 1 mslvpatnnyitpnlqkgttinnvgyvkvifkppylskgtdcscvvtldvqcnvklctcl 60
Qy 61 LFSGNVEALPIIYKNGDIVRHLKIOVYKKEGTGTSFGFASLTFTGLGAPTIPRTSS 120
Db 61 lfsngvealpIiykngdivrhlkiOvykKetqgItssgfAsltftgltgApIprtss 120
Qy 121 KYFNFTTDEHKMVEALRWASTHMSPSWTLKLCDDVQPMQYFDLTCLQLGKAEVDGASF 180
Db 121 kyfnfttdhkMvealrwasthmspswtlklcdvqpmqyfdltclqlgkaevdgastl 180
Qy 181 LKWDGTRTPPSPSRVLIQDLVLEGDLSHIRLQNLITDILVVDNHVHVARSLKVGSELR 240
Db 181 lkwdgtrtpspsrvliqdlvlegdshlrlqnlitdilvvdnhvhvarslkvgsflr 240
Qy 241 IYSLHTKLSQNSNQTLMSLEFHLHGCTSYGRGIRVLPESNSDVQDLKDLLESANLTAN 300
Db 241 IyslhtklSqnsnqTlmsleFhlhgctsygrgIrVlpesnsdvdlkdllesanltan 300
Qy 301 QHSDVICQSEPDSPSSGSVSLYEVRCCOQLSATILTDHQYLERTPLCALIKQKAPQOY 360
Db 301 qhsdvicqsepdspssgsvslyeverccoqlsatiltldhqylertplcalikqkapy 360
Qy 361 RIRAKLRSYKPRRLFQSVKLHCPKCHLLQEVPHGEGDLTIPODGATKTPDVKLQNTSLYD 420
Db 361 riraklrsykprrlfqsVklhCpkchllQevphgEGdlTIPOdgATkTPdvklQntslYd 420
Qy 421 SKIWTTKNQKGRKVAHVFKVKNNGILPLSNCELLLEGGTLSEICKLSNKFNSVIPVRSGH 480
Db 421 skiwtcknqgrkvavhfvknngilplsnecellleggtlseicklsnkfnsvipvrsg 480
Qy 481 EDLELDDLAPFLIOGTIIHHYGTGYCT 507
Db 481 edlelldsapfligtthhygckqcs 507
```

RESULT 3

AAB95120

ID AAB95120 standard; Protein: 634 AA.

XX AC AAB95120;

XX DT 26-JUN-2001 (first entry)

XX DE Human protein sequence SEQ ID NO: 17113.

XX KW human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX OS Homo sapiens.

XX PN EP1074617-A2.

XX PD 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-0116126.

XX PR 29-JUL-1999; 99JP-0248036.

XX PR 27-AUG-1999; 98JP-0300253.

XX PR 11-JAN-2000; 2000JP-0118776.

XX PR 02-MAY-2000; 2000JP-0183767.

XX PR 09-JUN-2000; 2000JP-0241899.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX DR WPI; 2001-318749/34.

XX PT Primer sets for synthesizing polynucleotides, particularly the 5602

XX PT full-length cDNAs defined in the specification, and for the detection

XX PT and/or diagnosis of the abnormality of the proteins encoded by the

XX PT full-length cDNAs .

XX

PS

Claim 8: SEQ ID 17113; 2537pp + CU KUM; English.

XX

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

SQ Sequence 634 AA;

Query Match 96.5%; Score 2637; DB 22; Length 634;

Best Local Similarity 99.2%; Pred. No. 9.5e-263;

Matches 503; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MSLVPATNVIYTPNLQKGGTIVNVYCVVKKFPFPPYLSKGTGYCVVTVTDQTNVLTCL 60

Db 1 mslvpatnviytpnlqkgttinnvgyvkvifkppylskgtdcscvvtldvqcnvklctcl 60

Qy 61 LFSGNVEALPIIYKNGDIVRHLKIOVYKKEGTGTSFGFASLTFTGLGAPTIPRTSS 120

Db 61 lfsngvealpIiykngdivrhlkiOvykKetqgItssgfAsltftgltgApIprtss 120

Qy 121 KYFNFTTDEHKMVEALRWASTHMSPSWTLKLCDDVQPMQYFDLTCLQLGKAEVDGASF 180

Db 121 kyfnfttdhkMvealrwasthmspswtlklcdvqpmqyfdltclqlgkaevdgastl 180

Qy 181 LKWDGTRTPPSPSRVLIQDLVLEGDLSHIRLQNLITDILVVDNHVHVARSLKVGSELR 240

Db 181 lkwdgtrtpspsrvliqdlvlegdshlrlqnlitdilvvdnhvhvarslkvgsflr 240

Qy 241 IYSLHTKLSQNSNQTLMSLEFHLHGCTSYGRGIRVLPESNSDVQDLKDLLESANLTAN 300

Db 241 IyslhtklSqnsnqTlmsleFhlhgctsygrgIrVlpesnsdvdlkdllesanltan 300

Qy 301 QHSDVICQSEPDSPSSGSVSLYEVRCCOQLSATILTDHQYLERTPLCALIKQKAPQOY 360

Db 301 qhsdvicqsepdspssgsvslyeverccoqlsatiltldhqylertplcalikqkapy 360

Qy 361 RIRAKLRSYKPRRLFQSVKLHCPKCHLLQEVPHGEGDLTIPODGATKTPDVKLQNTSLYD 420

Db 361 riraklrsykprrlfqsVklhCpkchllQevphgEGdlTIPOdgATkTPdvklQntslYd 420

Qy 421 SKIWTTKNQKGRKVAHVFKVKNNGILPLSNCELLLEGGTLSEICKLSNKFNSVIPVRSGH 480

Db 421 skiwtcknqgrkvavhfvknngilplsnecellleggtlseicklsnkfnsvipvrsg 480

Qy 481 EDLELDDLAPFLIOGTIIHHYGTGYCT 507

Db 481 edlelldsapfligtthhygckqcs 507

RESULT 4

AAB92742

ID AAB92742 standard; Protein: 634 AA.

XX AAB92742:
 AC 26-JUN-2001 (first entry)
 DT
 DE Human protein sequence SEQ ID NO:11188.
 XX
 KW Human; primer: detection; diagnosis; antisense therapy; gene therapy.
 XX Homo sapiens.
 OS
 XX EPI074617-A2.
 PN
 XX
 PD 07-FEB-2001.
 PF
 XX 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI: 2001-318749/34.
 XX
 PT primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs.
 PS
 PS Claim 8: SEQ ID 11188; 2537bp + CD ROM; English.
 XX
 XX The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialized methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human CDNA sequences; AA892446 to
 CC AA895893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 CC
 XX Sequence 634 AA:

Query Match 96.2%; Score 2628; DB 22: Length 634;
 Best Local Similarity 99.0%; Pred. No. 8.1e-262;
 Matches 502; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSIVPATNYTPTLNGKGTIVNVGVVFPFPPVLSKGTGCVSVTVTDQTNVLTCL 60
 DB 1 msivpatnytptlngkgtivnvgvvfpfppvlskgtgcvsvtlvdqtnvltcl 60
 QY 61 LFGSGNEALPIIYKNGDIVFHRKLIQVYKKEGTGSSGFASTREGTLGAPITPTSS 120
 DB 1 lfgsgnealpiiyrkngdivfhrklioqvyykkggtgssgfastregtlgapitptss 120

DB 61 llsngyealpiiyknodivrfhrkliqvykkggtgssgfastregtlgapiptss 120
 QY 121 KYNFTTTEDEHKWVEALKWASTHMSFSTLLKLCIDYQVPMQVFDLTCQLAGKAEVGCASFL 180
 DB 121 kynftttedehkwealrwasthmsfswllkicdvqpmqyfdltcqlagkaevgcasfl 180
 QY 181 LKVMDCGRTPFPSPRWVLIQDLVLEGDSLHRLQNTLIDILVYDNIHVARSLLKGSFPLK 240
 DB 181 lkvmdcgrtppspwrvliqdlvlegdshrlqntlidilvydnihvarsllkgsfll 240
 QY 241 TYSLHTRKLSMNSLQNTMLSLFPHLHCTSYGKCIHYVLHESNDVQLKDLISANLTAN 300
 DB 241 tyslhtklsmslqntmlslfphlhctsygkcihyvlhensdvqlkdlisnltan 300
 QY 301 OHSHVIVCOSERPDSPSSGSVSLYEVRKQQLSATIITTHOVLEKPTLCALIKOKAPQVY 360
 DB 301 ohshvivcoserpdspssgsvslyevrkqqlsatitthovlekpvlcailkokaqvy 360
 QY 361 RIRAKLSYKPRFLQSVKLIHCPKCHLLQEVPIEGDLDIIFDQAVKTVDPVKLQNTSLYD 420
 DB 361 riraklsykprrlfsvklihcpkchllqevpiegdlidiffdqavtktpvdklqntsl 420
 QY 421 SKIWTTKKQCKRKAIVHVKNNCILLQSNELLLKGTLSLSEICKLSNKPNSVIVKSGH 480
 DB 421 skiwttkkqckrkavhvknnclllqsnelllkgtslslseicklsnkpnsvipvsg 480
 QY 481 EDLELDDLSAPFLIQGTHHGYCTGCTP 507
 DB 481 edlelddlsapfliqgthhgyctgctp 507

RESULT 5
 AAM41911
 ID AAM41911 standard; protein: 307 AA.
 XX
 AC AAM41911:
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 6842.
 XX
 KW Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; hemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW Leukemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HSE-) HSEO INC.
 XX
 PI Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao Qa, Zhou P, Goodrich R, Drmanac R;
 DR WPI: 2001-442253/47.
 DR N-PSDB: AA161067.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 9, 2002, 20:29:26 ; Search time 3737 Seconds
(without alignments)
8409.721 Million cell updates/sec

Title: US-09-816-248-12

Perfect score: 1905
Sequence: 1 atgtcttggtccagcaac.....ttgcagaagatgaatactaa 1905

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.in.*

18: em.in.*

19: em.om.*

20: em.or.*

21: em.ov.*

22: em.pat.*

23: em.ph.*

24: em.pl.*

25: em.ro.*

26: em.sts.*

27: em.sy.*

28: em.in.*

29: em.vi.*

30: em.htgo_hum.*

31: em.htgo_inv.*

32: em.htgo_rod.*

33: em.htg_hum.*

34: em.htg_inv.*

35: em.htg_rod.*

36: em.htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB	ID	Description
1	1905	100.0	2631	9	AK001230	AK001230 Homo sapi
2	1901.8	99.8	2029	9	BC002923	BC002923 Homo sapi
3	1901.8	99.8	2077	9	AK001935	AK001935 Homo sapi
4	1901.8	99.8	2383	9	AK022580	AK022580 Homo sapi
5	1842.6	96.7	3192	9	AB066545	AB066545 Macaca fa
6	1268.4	66.6	1752	9	AK026234	AK026234 Homo sapi
7	408.4	21.4	2012	9	HSM800212	AL050120 Homo sapi
8	208.6	11.0	154959	9	AC004925	AC004925 Homo sapi
9	104.8	5.5	320	11	G12915	G12915 swSS2590 Er
10	51.2	2.7	1141	6	AX083744	AX083744 Sequence
11	45.8	2.4	127650	1	MPULM01	AL445563 Mycoplasma
12	45	2.4	110000	2	PFMAL13P2_1	Continuation (2 of
13	43.8	2.3	1141	6	AX083744	AX083744 Sequence
14	42.8	2.2	34641	3	DMBN32G13	AL035631 Drosophil
15	42	2.2	96310	9	AL136444	AL136444 Human DNA
16	42	2.2	159475	2	AC021378	AC021378 Homo sapi
17	42	2.2	169421	2	AC036163	AC036163 Homo sapi
18	42	2.2	172537	2	AC073122	AC073122 Homo sapi
19	42	2.2	174672	9	AC039057	AC039057 Homo sapi
20	42	2.2	182048	2	AC027439	AC027439 Homo sapi
21	42	2.2	182383	2	AC021745	AC021745 Homo sapi
22	41.6	2.2	181596	2	AC068158	AC068158 Homo sapi
23	41.4	2.2	60565	2	AC023852	AC023852 Homo sapi
24	41.4	2.2	153724	2	AC041034	AC041034 Homo sapi
25	41.4	2.2	209258	2	AC019234	AC019234 Homo sapi
26	41.2	2.2	40255	3	U41272	U41272 Caenorhabdi
27	41.2	2.2	110000	2	CEY39B6_1	Continuation (2 of
28	41.2	2.2	110000	2	CEY39B6_2	Continuation (3 of
29	41.2	2.2	154386	2	AC016314	AC016314 Homo sapi
30	41.2	2.2	158581	2	AL356485	AL356485 Homo sapi
31	41.2	2.2	180283	2	AL592406	AL592406 Homo sapi
32	41.2	2.2	207022	3	CEY39B6B	AL132896 Caenorhab
33	41.2	2.2	246237	3	CEY39B6A	AL132948 Caenorhab
34	41	2.2	7152	1	BA009185	U09185 Buchnera ap
35	40.8	2.1	39726	3	CER0411	292824 Caenorhabdi
36	40.8	2.1	150993	2	AC024289	AC024289 Homo sapi
37	40.8	2.1	195368	2	AC068399	AC068399 Homo sapi
38	40.8	2.1	198551	2	AF284562	AF284562 Homo sapi
39	40.6	2.1	124004	9	AC004947	AC004947 Homo sapi
40	40.6	2.1	181561	2	AC015911	AC015911 Homo sapi
41	40.4	2.1	2759	3	MILTMCX1	X02438 Leishmania
42	40.4	2.1	20992	3	LEKRWAX	M10126 Leishmania
43	40.4	2.1	123159	2	AC091880	AC091880 Homo sapi
44	40.4	2.1	144598	2	AC092946	AC092946 Homo sapi
45	40.4	2.1	146602	2	AC064795	AC064795 Homo sapi

ALIGNMENTS

RESULT 1

AK001230

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (sites)

Isopai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,

Nishikawa, T., Nagai, K., Sugano, S., Shiratori, A., Sufo, H.,

Wajutsuna, M., Hosoi, T., Kaku, Y., Kodaira, H., Kondo, H.,

Sugawara, M., Takahashi, M., Chiba, Y., Ishida, S., Murakawa, K.,

Ono, Y., Takiuchi, S., Watanabe, S., Kimura, K., Murakami, K.,

Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A.,

Nakamura, Y., Nagahara, K., Masuho, Y., Ninomiya, K. and Iwayanagi, T.

AK001230 Homo sapiens cDNA FLJ10368 fis, clone NT2RM2001544. 22-FEB-2000

AK001230 Homo sapiens cDNA FLJ10368 fis, clone NT2RM2001544.

AK001230.1 GI:7022354

oligo capping: fis (full insert sequence).

Homo sapiens teratocarcinoma cell line: NT2 cDNA to mRNA,

clone: lib-NT2RM2 clone: NT2RM2001544.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished (2000)
REFERENCE 2 (bases 1 to 2631)
AUTHORS Isegaki, T. and Otsuki, T.
JOURNAL Direct Submission
Submitted (16-FEB-2000) to the DDBJ/EMBL/Genbank databases. Takao Isegaki, Helix Research Institute, Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3951, Fax: 81-438-52-3952)
COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan. cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute of Virology, Institute of Medical Science, University of Tokyo.
FEATURES
source
1. .2631
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="NT2"
/cell_type="Teratocarcinoma"
/clone="NT2RM2001544"
/clone_1id="NT2RM2"
/note="cloning vector: pME18SFL3; mRNA from uninduced NT2 neuronal precursor cells."
24. .1928
/note="unnamed protein product"
/codon_start=1
/protein_id="BAA91568.1"
/db_xref="GI:7022355"
/translation="MSI VPATNYITPLNOLKGTIVVNYGVVKKFKPYLSKGTDFC SVYTIADQTNVTKLFCCTNYEALPIYKNGDIYVPHRLKIOYKKKEPGITSSGFA SLPEEGIGAPDIIIPRTSSKFNFTEDHMMVEALRWMASTHNSPSMTLLKLDYOMO YPDLTGCLICKAFVKSFLPKYKMDCTRTPTSMRYLTODVIEGHLSTHRLQNTT DILVYDHYVAVSLKVGSLKYSLSHLKLSKNSNLSNLMLSLEHLKICSTYKGRJIK WLPESNDVDQLKDLSEANLIVANQHSVDVTCSEPDSPSSGVSLSLVEVERCQDLDA TLTLDHYLERPLCAILKOKAPOYRIRAKLSYKPRILFQSVKLHCKPKHLLQEPV HEGEDLDIIFQDGAKTTPVVKLQNTSLSDSKJMTTKXOKGRKAVHVKNGILPLSNE CLLIEGRTLSKIKLSKNSKFNYSIVPVSCHDELELDLSAPLIGCTIHHYCKOCS LRSIQNLNSLVKTSMPSSVALELCIVLYVYVWTFPLDQGTGLEAVYLMDSKPY QIPASEVLMDDLOKSVDMIMDFCPGKIDVYPLWLEFISINVTNGTDMDQICQI FDTTVAEDVY"
BASE COUNT 826 a 473 c 491 g 841 t
ORIGIN
Query Match 100.0%; Score 1905; DB 9; Length 2631;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1905; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 324 TTTCATCTTTGACGTTTGAGGCAACTTGGAGCCCTATCATCTGACCTTCAC 383
Oy 361 aagtaatttaacttaacttaacttaaggaccacaaatgtagaaaccccttaagtctggaga 420
Db 384 AAGTATTTTAACTTCACTACTGAGGACACAAAATGGTAGAACCTTACGCTTGGGCA 443
Oy 421 tctactcaatgtcacccgctctggagacatlaaaaaatlgatgagatllcaaccatlgag 480
Db 444 TCTACTCATATGTCACCGCTCTTGACATTCTAATAATGTGTATGTTCCAGCCATTCCAG 503
Oy 481 tatttgaacctgacttgtaagctcttgggcaagaagagggagagacatcattctt 540
Db 504 TATTTTCACTGACTGTGTGAGCTTGTGGCAAGCAAAATGGACGACCATCATTTCTT 563
Oy 541 ctlaaaagtatggatagacacagacacatctcaatcttgagagctctaaacaagac 600
Db 564 CTAAACCTATGGCATGGCACACGACACCACTTTCATCTTCCAGACTTTATACAGAC 623
Oy 601 ctctgctcttgagagatgagatlllaagatctgcatctgagctctacaaatctgagaaacat 660
Db 624 CTCTCTCTTCAGACCTGATTAATCACTCAATCCATCGGCTCAAAATCTCAATACAT 683
Oy 661 ttagctacgataacagatgctcagtgagcaagatctctgaagatlgagagctctctaga 720
Db 684 TTAGCTCAGCATTAACCATGTTTCATGTGGCAAGATCTGTGACGTTTGAGAGCTTCTTGA 743
Oy 721 atctatagctctcaatcacaaactctcaatcaatgaaatcagagaaatcagaaatgtaagt 780
Db 744 ATCTATTAACCTTCATACGAAACCTCAATCAATCAATCAATCAATCAATCAATCAAT 803
Oy 781 ttagagatctcaatctcaatcaagatgaaacagcttgaagatcagagatctctgcaaga 840
Db 804 TTGACATTTTCATCTTCATGAGAGTACCATTTACGGTGGGGAATCAGGCTTGGCAGAA 863
Oy 841 aatlaactctgaatgtaagatcaaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaat 900
Db 864 ACTAACTCTCATCTGATCACTGATCACTGATCACTGATCACTGATCACTGATCACTGAT 923
Oy 901 cagatctcagagatgctcctgcaatcagagacatcagagacatctcgaagctcagatca 960
Db 924 CAGCATTCACATCTTATCTCTCATCATCAGAACCTGACGACAGCTTTCAGAGCTTGGATCA 983
Oy 961 gtaacatlaacagagatgaaagaaatgaaacagctctacatctgcaacatctcaagatcat 1020
Db 984 GTATCATTTATACAGAGGTACAAAGATGTCAACAGCTATCTGCTACATCTTACAGATCAT 1043
Oy 1021 cagatctcagagagagacacactatctgcatcttgaagaacaaagacccctcaacaatcac 1080
Db 1044 CAGTATTTTGGAGAGGACACCACTATGTGCAATTTTGAACAAAAAGCTCTCAACAAATAC 1103
Oy 1081 cgcattccgagcaaaatgagatcatalaagccagaaagactatctcaatctctgtaaaact 1140
Db 1104 CCCATCCGAGCAAAATTTGAGGTCAATTAAGCCCAAGAACATTTTCAGCTGTTAAACTT 1163
Oy 1141 catctccctaaatgtaatctgctcgaagaagaatctcagatcagagagatcttgatataact 1200
Db 1164 CATTCGCCCTAAATGTCAATTTGCTGCAAGAAAGTCCACATGAGGGGATTTGGATTAATT 1223
Oy 1201 tttcagagatgagcaacttaaaaccccaatctgtaagatctacaaaaatacatcatatagat 1260
Db 1224 TTTCAAGATGTGCAAACTVMAAACCCACAGTTGCAAGTTTACAAATATCATTTATATGAT 1283
Oy 1261 tcaaaaatcttgagacacatlaaaatcgaagaagagaaagatagagatctcatcttgtagaa 1320
Db 1284 TCAAAATCTTGACACACTVMAAATCAAAAAGACGAAAGATGACAGTTCAATTTGTGAAA 1343
Oy 1321 aatlaatgtaattccgcctcttcaaaatgaaatgtaatcttgaatagaagagagatcacctc 1380
Db 1344 AATATATGATTTCTCCCTCTTCAATATGAATGTCTACTTTGATAGAAAGAGATCACTC 1403
Oy 1381 agtgaattctgcaaacctctcgaaacagatctaaatagtgtaattctctgtgagatctcgccac 1440
Db 1404 ACTGAATTTTGCAAACTCTCCAAACAAATTTAATAGTTAATTTCTGTGACATCTGCCAC 1463

HEGDLIIIFODGATKTPDVVKLQNTSLYDSKIWTTKNQKGRKVAHFVKNIGLPLSNE
CLLIIEGGTISEICKLSNKFNSVPIVRSHEDLELLDLSAPFLIOGTIIHYGCKQCSS
LRSTONLSLVKTSWIPSSVAELGIVPQYVEMTFTLDDGTGVLEAYLMDSDKFF
QIPASEVLMDDDLQKSDVMDIMDFCPPIKIDAPWLECFIKSYNTVNGTDNOCYQI
FDTIVASDVI*
BASE COUNT 657 a 401 c 393 g 626 t
ORIGIN

Query Match		99.8%	Score 1901.8	DB 9	Length 2077	
Best Local Similarity		99.9%	Pred. No. 0			
Matches 1903		Conservative 0	Mismatches 2	Indels 0	Gaps 0	
QY	1	atgtcttttgggtccagcaacaattatatacacacccctgaatcaacttaagggtggt	60			
Db	131	ATGTCTTTGGTTCAGCAACAATATATATACACCCCTGAATCAACTTAAGGGTGT	190			
QY	61	acaattgcaatgtctatggtgtggaagtcttttaagcccccatatctaagcaagga	120			
Db	191	ACAAATTCATGTCATGGTGTGTGAAGTCTTTAAGCCCCCATATCTTAAGCAAGGA	250			
QY	121	actgattatgtctagtgtaactattgtgacagacaatgttaaaactaaacttgctg	180			
Db	251	ACTGATTATGCTCAGTGTAACTATTGTGGACCCAGACAATGTAAACTTAAC	310			
QY	181	ctcttttagtgaactatgaagcccttccataatttataaaatggaatattgttcgc	240			
Db	311	CTCTTTTAGTGGAACTATGAAGCCCTTCCAATAATTTATAAAATGGAGATATT	370			
QY	241	ttcacagactgaagatcaagtatataaaaggagactcaggtatataccacagctctgc	300			
Db	371	TTTCACAGGCTGAAGATTCAGTATATAAAGGAGACTCAGGGTATCACCAGCTCGC	430			
QY	301	tttgcatctttgacgtttgaggaactttgggaagccctatcatatcctgcactcaagc	360			
Db	431	TTTGCATCTTTGACGCTTTGAGGAACTTTGGGAGCCCTATCATACCTCGCACTCAAGC	490			
QY	361	aagtattttaacttcaactaggaagacacaaatgtagaagcttcagtggtttgggca	420			
Db	491	AAGTATTTTMACTTCACTACTAGGAGCCCAAAATGGTAGAAGCTTACGTTCTTGGCA	550			
QY	421	tctactcatatgcaacgcttcttgacatatactaaaattgtgtgatttcaagccaatgcag	480			
Db	551	TCTACTCATATGTCACGCTCTTGGACATTATCTAAATTTGCTGATGTTCAAGCAATGCAG	610			
QY	481	tattttgacactgactgttcagctcttgggaagcaaaatgtagaagcagctattctctt	540			
Db	611	TATTTGACCTGACTTTGTGAGCTCTTTGGGCAAAAGCAGAAAGTGGAGCAATCATTTCTT	670			
QY	541	ctaaaggtatgggaaggacacagacacatttccattttgagaggtctttaaatacaagac	600			
Db	671	CTAAAGGTATGGGATGGCAGCAGGACACCAATTTCCATCTTCGACAGTCTTTAATACAGAC	710			
QY	601	cttgcttctgaaggtgatttaagtcaactcaactcaggtctcaaaaatctggaataatgacatt	660			
Db	731	CTTGTTCTTTGAAGGTGATTAAAGTCACATCCATCGGCTACAAAATCTGCACAAATAGACATT	790			
QY	661	ttagtctcaagataaacatatttcatgtggcaagatctcttgaaggtcttgaagcttcttga	720			
Db	791	TTAGTCTACGATAAACATGTTCTATGCGCAAGATCTCTGAAAGTTGGAGGCTTCTTTAGA	850			
QY	721	atctatagccttcaataccaaacttcaatcaatgaattcagagaatcagagaatgttaagt	780			
Db	851	ATCATATAGGCTTTATACCAAACTTCAATCAATTCAGAAATTCAGAAATGTTAAGT	910			
QY	781	ttaaggtttcaacttctatggaaggtaccagtttacggtcgggaatacaggtcttgccagaa	840			
Db	911	TTAGAGTTTTCATCTTCATGCGAGTACCAGTTACGGTCCGGGAATCAGGGTCTTCCAGAA	970			
QY	841	agtaactctgattgtgtatcaactgaaaaaggaatttagaattctgcataatttgacagccaat	900			
Db	971	AGTAACCTGATGTGATCACTCACTGAAAAAGGATTTAGAATCTGCAAAATTTGACAGCCAAAT	1030			

RESULT 4
AK022580
LOCUS
DEFINITION

2383 bp mRNA
Homo sapiens cDNA FLJ12518 fis, clone NT2RM2001805.
PRI 29-SEP-2000

QY	901	cagcattcagatgtattctctgtcactcagaacctgacgacagctttccagcctctggaatca	960
Db	1031	CAGCATTTCAGATGTTATCTGCTCAATCAGAACCTTCAGCAGACAGCTTTTCCAGCTTGGATCA	1090
QY	961	gtatcattatcagaggtagaagaatgtcaacagctatctgctacaatacttacagatcat	1020
Db	1091	GTATCATTTATACGAGGTAGAAAGATGTCAACAGCTATCTGCTACAAATACTTACAGATCAT	1150
QY	1021	caattattggagagagaccaccattatgtgcattttttaaacaacaaagctcctcaacaatc	1080
Db	1151	CAGTATTTGGAGAGGACACACACTATCTGCCTATTTTCAACAAAAAGCTCTCTCAACAATAC	1210
QY	1081	cgaatcgaagcaaaaatggaagtcataataagcccagagaagactatttcagctctgttaaacctt	1140
Db	1211	CGCATCCGAGCAAAATTTGAGTTCATATAAGCCACAGAACTATTTTCAGTCTGTGTAACCTT	1270
QY	1141	catgcccctaattgtcattttgctgcgaaggttccacacagagggcatttggataaatt	1200
Db	1271	CATTGCCCTAAATGTCATTTCTTGCAGAAAGTTTCCACATGAGGGCGATTGGATATAATT	1330
QY	1201	tttcaggatggtgcaactaaaacccagttgtcnaagttacaacatacatcattatgat	1260
Db	1331	TTTTCAGGATGCTGCAACTTAAACCCGAGATGTCAGCTACAAAATACATCATTTATATGAT	1390
QY	1261	tcaaaaatctggaccactaaaatcaaaaggacgaaaagtagcaggttcatttttggaaa	1320
Db	1391	TCAAAATCTGGACCACCTAAAAATCAAAAGGACGAAAGTAGCAGCTTCATTTTGTGAAA	1450
QY	1321	ataaattgttatctctgcgctttcnaatgaatgtctacttttggatgaagaggtacactc	1380
Db	1451	AATAATGGTATTTCTCCGCTTTTCAAAATGAATGTCTACTTTTGATAGAAAGAGGTAGACTC	1510
QY	1381	agtgaaattctgcaactctcgaacaaagtctaaagtgatgtaattctctgtgagatctggccac	1440
Db	1511	AGTGAATTTGCAAACTCTCGAAACAAGTTTAAATAGTGAATTCCTGTGAGATCTGGCCAC	1570
QY	1441	gaagacctggaactttttggaccttttcagcaacctttcttatatacaaaagaaacaaatcacatcac	1500
Db	1571	GAAGACCTGCAACTTTTGGACCTTTTCAGCACCACTTTCTTATACAAAGCAACAATACATCAC	1630
QY	1501	tatggatgtaaacagtgattctatgtttggaatccatatacaaaaatcctaataatccctcgttgtat	1560
Db	1631	TATGGATGTAACAGTGTTCTAGTTTCAGATCCCATACAAATCTAAATTCCTGTGAGATCTGGTGTAT	1690
QY	1561	aaacacatcgtggaactctctctctgtggaagagcctgggtattctacacctcccaatat	1620
Db	1691	AAACATCTGTTGGATTCCTTCTCTGTCGAGAAGCACCTGGGTATTTCTACCCCTCCAATAT	1750
QY	1621	gtcttcttctatgacctttcacacttgatggaacagggagtagctagaagaacctctcctcatg	1680
Db	1751	GTGCTTTGTTATGACCTTTTACACTTTGATGATGGAAACAGGAGTACTAGAACCTTATCTCATG	1810
QY	1681	gattctgacaataattctccagatctcagagcatcagaagttcttgaatggatgtagaccttcag	1740
Db	1811	CATTCTGCACAAATTTCTCCAGATTCAGATCAGAAAGTTCTCATGATGATGATGACCTTCAG	1870
QY	1741	aaatgttgtatgatcatgtgatatattttgttctctcaggaataaaaaatctgagatcat	1800
Db	1871	AAAAGTGTGATATGATCATGATATGTTTGTCTCTCCAGCAATAAAAAATTTGATGGATAT	1930
QY	1801	ccgtgggtggaatgcttctcacaagtcatacaaaatggacaaatggacacagataatcaaat	1860
Db	1931	CCGTGGTTGGATTCCTTCATCAAGTATACAAATGTCACAAATGGAACAGACAGATAATCAAAAT	1990
QY	1861	tgtatcagatattttgacaccacagttggagaagatgtaatactcaa	1905
Db	1991	TGCTATCAGATTTTTCAGACCCACAGTTTGCAGAAAGATGTAATCTAA	2035

ACCESSION AK022580.1 GI:10434050
VERSION AK022580.1 GI:10434050
KEYWORDS Oligo capping; fis (fuj) insert sequence;
SOURCE Homo sapiens tefalocarcinoma cell_line:NT2 cDNA to mRNA,
clone_11b:NT2RM2 clone:NT2RM2001805.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (sites)
AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y.,
Matsunawa,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,
Wakamatsu,A., Nakamura,Y., Nagahara,K., Masuko,Y. and Sasaki,N.
NEDO human cDNA sequencing project
TITLE Unpublished (2000)
JOURNAL Unpublished (2000)
REFERENCE 2 (bases 1 to 2383)
AUTHORS Isogai,T. and Otsuki,T.
JOURNAL Direct Submission
TITLE Submitted (23-AUG-2000) to the DDBJ/EMBL/Genbank databases. Takao
Isogai, Helix Research Institute, Genomics Laboratory: 1532-3 Yama,
Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan: cDNA full insert
sequencing. Research Association for Biotechnology: cDNA library
construction, 5', 3'-end one pass sequencing and clone selection:
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
Source location/Qualifiers
1..2383
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="NT2"
/cell_type="tefalocarcinoma"
/clone="NT2RM2001805"
/clone_11b="NT2RM2"
/note="cloning vector: PMEB8FL3-mRNA from uninduced NT2
neutonal precursor cells."
439..2343
/note="unlabeled protein product"
/codon_start=1
/protein_id="BAB14110.1"
/db_xref="gi:10434051"
/translation="MSLVPAIVNYITPLNOLKGTIVVNVGVKFKPKPYLSKGTDC
SVYTIADOTNWKICLLPSCNVEALPIYKMGDIYRFRILIOYKKEQITSSGFA
SLTEGCTGAPJIIPTSSKYPNFTPEHKNWEALRYVASTHMSMTLLKIDYOPMO
YFDLTCOLLGAEDVGAFLKLVMDGTTPPSNRVYLQDLYEGDLSHTRDLONTI
DLYVDNHVHARSLKVGSLKRTISLHFKLOSMSSENOTMLSEPHLHGTSYGRIR
VLPESSNDVDOLKDESANLTANOHSOVTCQSEPDSPSSGVSLHVEERCOQLA
TILTHOYLERTPLCAILKOKAPOOYRIIRAKLRSYKPRRLQSVKLYHCPCHLQEVF
HEGDLDIIFODGATKTPVYKLONTSLDYSKIMTITNOKRGVAHFVNNQILPLSME
CILLTEGTLSEICLKSFNPSVLPVRSKHEDLELLDSAPFLQGTIHRGKQCS
LRSIONLSLVDTKSMIPSSVAEALGTYPLQYVFMPTLDDGVEALYLMDSKFF
QIPASEVLMDDDLQKSVDMIMDMCPPIKIDATPMLCFTKISYNTNGTNOICYOI
FTTYAEVPI"

Db 559 ACTGATTATTGCTCAGCTTCTAATCTATGTCAGCAGACAAATGTAATCTAATCTGCTG 618
Oy ctcctttagtggaactatgaagaagcccttcacaacttataaagaatgagatgctcgc 240
Db 619 CTCCTTAGTGGAACACTATGACCCCTTCATATTTATTAATAATGAGATATTGTTCCG 678
Oy 241 ttcacagaagcgaagatcgaagatataaagaagaagacgaagatgatacgaagctcggc 300
Db 679 TTTTCACAGCCTGACATATGACCTTATATTAATAAGAGACCTGACCTATTCACAGCTTCCG 738
Oy 301 ttgcacatcttgaagcgttgaaggaagccttggagagcccttcatacactcgaactcgaag 360
Db 739 TTTTCACATCTTGAAGCTTTGAGGAATATTGGAGAGCCCTCATATACCTTCGACATTCAGC 798
Oy 361 aagatatttgaactcgaactcgaagacgaagaagatgataagcgaagcgaagcgaagcga 420
Db 799 AGCTATTATTAACTTACATCTGACGACGACCAAAATGCTAGAAAGCTTACTGTTGGGCA 858
Oy 421 tctactcgaatgacagcgttgaagacatcttgaagacatcttgaagacatcttgaagac 480
Db 859 TCTACGTCATATGTCACCTGCTTCGACATTAATAATTCGTGATGTTGACCCATATGAC 918
Oy 481 tatlgaactgaactgacgctccttgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 540
Db 919 TATTGTCACCTGACCTTCTGACCTTGGGCAAGAGAGAGCTGACGACATCTTCTT 978
Oy 541 ctgaagataggaatgacagcgaagacgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 600
Db 979 CTAAAGCTATGCGATGCGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1038
Oy 601 ctgctccttgaagcgaagatllgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 660
Db 1039 CTGCTCTTTCAGCTGATTTAACTGACATTCGACATTCGCTGACATTCGACATTCGACATTC 1098
Oy 661 ttagctcgaagataggaatgacgctccttgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 720
Db 1099 TTAGCTTACCATTAACCATGTTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1158
Oy 721 atcctatagcctcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 780
Db 1159 ATCTATGCTGCTTCTATACCAAACTTCAATCAATCAATCAATCAATCAATCAATCAATCAAT 1218
Oy 781 ttagagctcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 840
Db 1219 TTAGAGCTTATGCTTCTTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1278
Oy 841 agtaactcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 900
Db 1279 AGTAACTGCTGATGCTGATCAACGAAAGAGATTTAGAAATTCGAAATTTGACAGCCAAAT 1338
Oy 901 cagcattcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 960
Db 1339 CAGCATTCAATGCTTATCTCTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 1398
Oy 961 gtaacatataagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 1020
Db 1399 GTATCATTTATACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1458
Oy 1021 cagatattggaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 1080
Db 1459 CAGTATTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1518
Oy 1081 cgaatcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 1140
Db 1519 CGCATCCGACCAAAATTTGAGCTCATATAGCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1578
Oy 1141 catgacctgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 1200
Db 1579 CATGCGCTTAATGCTATTTGCTGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1638
Oy 1201 ttcagagatgagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 1260

Db	1639	TTTCAGGATGTTGCACTAAACCCAGAGTGTCAAGCTACAAATACATCATATATGAT	1698
QY	1261	tcacaaatctgaccactaaaatacaaaagacgaagtagcaggttcattttgtgaaa	1320
Db	1699	TCAAAATCTGGACCACATAAAATCAAAAGGACGAAGTAGCAGTTCAATTTGTGAAA	1758
QY	1321	aataatggtattctcccgctttcacaatgaatgctactctttgatagagaggtacactc	1380
Db	1759	AATAATGTAATTCGCCGCTTTCAATCAATGCTCTACTTTTGATAGAGGAGGTACACTC	1818
QY	1381	agtgaatttcacaaactcgaacaagttaataagtgtaattcctctgagatgtggccac	1440
Db	1819	AGTGAATTTTGCMAACTCTCGAACAAGTTTAAATAGTGAATTCCTGTGAGATCTGGCCAC	1878
QY	1441	gaagacctggaaacttttgaccttccagcaccattctctatacaagaacaatacatcac	1500
Db	1879	GAAGACCTTGGAACTTTTGACCTTTTCAGCCACTTTCTTATACAGGAACATACATCAC	1938
QY	1501	tatgatgtcaaacagtggtcttagttgagatccatacaaaaatctaaatccctgggtgat	1560
Db	1939	TATGATGTAAACAGTGTCTAGTTTGAGATCCATACAAATCTAAATTCCTGTTGAT	1998
QY	1561	aaacatcgtgattcctctctctgtgagcagaagcactgggtattgtaccctcccaatat	1620
Db	1999	AAACATCTGTTGATTCCTTCTTCTGTGCGAAGCACTGGGTATTGTACCCCTCCAATAT	2058
QY	1621	gtgtttgttatgacctttcacacttgatggaaacagagtagtactagaagcctatctcatg	1680
Db	2059	GTGTTGTTATGACCTTTACACTTGATGATGGAACAGGAGTACTAGACCTATCTCATG	2118
QY	1681	gattctgacaaattcttccagatctccagatccagaagctcagatgagatgacattcag	1740
Db	2119	CATTCTGACAAATTTCTCCAGATTCACAGTTCAGAGTTCAGAGTCTGTGATGATGACCTTCAG	2178
QY	1741	aaaagtgtgatatgacatgatgatattttgtctccacaggaataaaaactgaatgat	1800
Db	2179	ANAGTGTGATATGATCATGATGATATTTTGCTCCAGGAATAAAATTTGATGATAT	2238
QY	1801	ccgtgtgtggaatgcttcacatcaatgacaaatgacaaatggacacataatcaaat	1860
Db	2239	CCGTGTGTGAATGCTTCATCACTCACTCAATGTCACAAATGGAACAGATCAATCAATTT	2298
QY	1861	tgetatcagatttttgacacacacagttgcgaagatgataactcaa	1905
Db	2299	TCCTATCAGATTTTTCACACCACACTTTCGACAGATGTAATCTAA	2343
RESULT	5		
LOCUS	AB066545	3192 bp	mRNA
DEFINITION	Macaca fascicularis brain cDNA clone: QtrA-10940, full insert sequence.	PK1	26-JUL-2001
ACCESSION	AB066545		
VERSION	AB066545.1	GI:15021899	
KEYWORDS	oligo capping; fls (full insert sequence).		
SOURCE	Macaca fascicularis adult male temporal lobe right cDNA to mRNA, clone: lib:macaque brain cDNA library QtrA-10940.		
ORGANISM	Macaca fascicularis		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae; Cercopitheciinae; Macaca.		
REFERENCE	1 (bases 1 to 3192)		
AUTHORS	Osada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hirai, M., Terao, K., Suzuki, Y., Sugano, S. and Hashimoto, K.		
TITLE	Isolation of full-length cDNA clones from macaque brain cDNA libraries		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 3192)		
AUTHORS	Hashimoto, K., Osada, N., Hida, M., Kusuda, J. and Sugano, S.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-JUL-2001) Katsuyuki Hashimoto, National Institute of Infectious Diseases, Division of Genetic Resources; 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan		

(E-mail: khashi@nih.go.jp, URL: http://www.nih.go.jp/yoken/genebank/, Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)	
Lab host:	TOP10
Vector:	pME18S-FL3 (Acc.No. AB009864)
R. Site1:	DraIII (CACTGTCTG)
R. Site2:	DraIII (CAGCATGTG)
Description:	1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCTTTTCTTTTCTTTT], double-stranded cDNA was synthesized using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of pME18S-FL3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Libraries were constructed by Sugano et al. (University of Tokyo, Institute of Medical Science). Custom primer used for sequencing
{ 5' end primer [CTTCTGCTCTAANAACCTGCG];	
3' end primer [CGACCTGCACCTCGACACA]):	
Location/Qualifiers	
1. .3192	
/organism="Macaca fascicularis"	
/db_xref="taxon:9541"	
/clone="QtrA-10940"	
/sex="male"	
/tissue_type="temporal lobe right"	
/clone_lib="macaque brain cDNA library QtrA"	
/dev_stage="adult"	
569. -2473	
/codon_start=1	
/product="hypothetical protein"	
/protein_id="BAB62219.1"	
/db_xref="GI:15021899"	
/translation="MSLVPATNYIYPLNQLKGGTIVNVYGVVKKFPYLSKGTDYC	
SVWTVDGTVNVTLLFSGNYEALPIYKNGDIVRFHRLKIOYVKKTGTTSSGFA	
SITFEETLGAPIIPRTSSKYFNFTEDHKHVTLRVMASTHMSPDWTLKLCVDQPMQ	
YDFLTCOLLGRAEVDGAFLLKVPDTRTPPSRWLVLDLVLEGSLSHRLQNLTI	
DILVTDNHVHVARSKVGSFLRIYSUHTKLQSMNSNQTMLSLEPHLGGTSTYGRIR	
VLPTNSVDQLKTLKESANLTANHISDVICQSEPDSPSSGVSLSLEYVERCOOLSA	
TILTDIYLIERTLCAILKQAPQYRIAKLSYKPRFLFSVKLHCPKCHLQLEVP	
REGDLDLIDQAGATKPDVKLQNTALYDSKWTTRKNQKRVAVHFNKNGLPLISNE	
CLLLLEGTLSIEICLKNKNSVIVPVSCHEDLELLDSAPFLQCTTHHYCKOCSS	
LRSTONLSLVDKTSKIPSSVAEVLGIVPLQYVFMVFTLLDQGTGVLAYLMDSDKFF	
QIPASEVLMDUULQKSMDMIMDMETPPGLIKIDAYIPWLECFIKSYNVTNLTQYQI	
PDITVAEDVI"	
BASE COUNT	1005 a 564 c 596 q 1027 t
ORIGIN	
Query Match	96.7%; Score 1842.6; DB 9; Length 3192;
Best Local Similarity	98.0%; Pred. No. 0;
Matches 1866; Conservative	0; Mismatches 39; Indels 0; Gaps 0;
QY	1 atgtcttttgcgaacaaatttatataataacacctgaatcaacttaagggtggt 60
Db	569 ATGCTCTTGTCTCCAGCACAAATTTATATATATACACCCCTGANTCACTTAAGGCTGCT 528
QY	61 acaattgtcaatgtctatgtgttbtgaaagttctttaaagcccccacatatctaagcaaaagga 120
Db	629 ACAATTTGTCATATGCTACCGTCTTGTGAAGTCTCTTTAAGCCCCCATATCTAAGCAAAAGCA 688
QY	121 actgaattattctcactgttgaactattgtgacagacagacaaatgtaaaactaactgcctg 180
Db	689 ACTGATATTATGCTAGTTGTAACAATTTGGGACCCAGACAAATGTATAAATTAACCTTGCCTG 748
QY	181 cctcttagtgaacactgaagcccttccaaataattataaaaaaggagatattgttcgc 240
Db	749 CTCCTTAGTGGNAACATATGAACCCCTTCCAATAATTTATAAATGGAGATATTCTTCGC 808
QY	241 ttccacaggtcgaagattcaagtataataaaaggagactcaggttatcaccagctctggc 300
Db	809 TTTCACAGCTCAACATTCACATATATAAAAGGAGACTCAGGGTATCACCAGCTCTGCG 868
QY	301 ttgcatctttgacgtttgagggaaactttgggagcccttatcatcaccctgcacttcaagc 360

Db	869	TTTGGATCTTTGACGTTTGAGGAACTTTGGAGGCCCTTATACCTGGACATTCAAGC	928
Oy	361	aagatatttbaacttcaactaactgaagaccacaataatgtagaagcccttaagtlttgggca	420
Db	929	ANGTATTTTAACTTCACTACTGAGGACCACAAAATGGTAGAAACTTTCAGCTGTGTTGGGA	988
Oy	421	tcctacataatgacccgctcttgagacatactaaatltgtatgtlctcaagccatagcg	480
Db	989	TCTACTATATGTCACCGCTTGTGGACATTACTAAATATTGTGTGATGGCCCAATGCGAG	1048
Oy	481	tatttgaacctgaacttgaqclcttggcdaagcgagatgagccgagatcaatlctt	540
Db	1045	TATTTTGAACCTGCATCTGTCAGCTCTTTGGGCAACAGAAAGTCGATGAGCATATTTTCTT	1108
Oy	541	ctaaaggtatgggaatgacacagagacacatlttcacatcttgaatgaqcltlaalacaagc	600
Db	1109	CTAAGATATGGATGGGACACAGACCAATTTCCATCTTGGAGACGCTTAAATACAAAGC	1168
Oy	601	cttgltcttgagagtgatltaaagtcacatccatcgagctgacaaatctcgacatataagcat	660
Db	1169	CTTGTCTTGAAGGTGATTTAAATCATCAATCCATCGGGCTACAAAATCTGCATATAGCATTT	1228
Oy	661	ttaqctacagataaacatgtlcatltgtagaagaatcltgaagatltgaaagcttcttga	720
Db	1229	TTAGCTACACATTAACCATGTTTCATGTGGCAAAATCTCGAAAGTTGGAACTTTCTTTAGA	1288
Oy	721	atctatagcccttcaaccgaacttcaataaagaatctagaagaatctgaacaaatlttaagt	780
Db	1289	ACTCTTACCCTTCATACCAAACTTCATATCATATTCATGAAATTCAGAAATTCAGCAATGTTAAGT	1348
Oy	781	ttagagcttcaactcttcaaggagtaacagltacagltcagtguaatcagtglttggcagaa	840
Db	1349	TTAGAGTTTCATCTTCATAGGAGCTACACAGTTATGTGGGGAAATCAAGGCTTGTCCAGAA	1408
Oy	841	aatbaacttgaatolagaataacttgaagaagaatltagaatcltgaabaatltaacagccaat	900
Db	1409	ACTAACTCTGATGTGATCAACTGAAAAAAGATTTAGATCTGCAAAATTTGACAGCCAAAT	1468
Oy	901	cgagcatgaagatgttalcctgtcaactcagaaaccttgaagacagcttccaaagctctgagta	960
Db	1469	CACCATTTGAGATGTTATCTCTCATATCGAAACCTGATGAGACAGCTTTCCAAAGCTGTGATCA	1528
Oy	961	gtatcattacacagagtgaagaatgttcaaacagctatcltqtacaaatacttacaagcat	1020
Db	1529	GTTATCTTCTATAGAGTAAAGAAAGATGTCAACACAGCTATCTGTACAAATATCTTACGATCAT	1588
Oy	1021	cagatattggaaggaacacacataatgtgcattltgaacaaabaagctcctcaacaatac	1080
Db	1589	CAATATTTTGGAGACACACACTGTGGCAATTTTGAAACAAAAGCTCTCTCAACAAATAT	1648
Oy	1081	cgacatccgagcaaaatltgaatgataaagccagaagacatlttcagltcgttaact	1140
Db	1649	CGCATCCGAGCAAAATTAGTGATTAATTAACCCAGAAAGATTAATTAAGCTGTGTAACCTT	1708
Oy	1141	catlccccaatagtcatlcttgctgaagaaggttcccaatgaagggcgaattgtgataaat	1200
Db	1709	CATTGCCCTAAATGTCATTTACTGTCAAAAGATTTCACGTGAGGTCATTTGGATTAATTT	1768
Oy	1201	lttcaagatgtgtgcaactaaaacccacagltgtcaaglttacaabaatcacatclataagat	1260
Db	1769	TTGCAGAGATGTGCAAAATTAACCCAGATGTCAAAATATACACATTATATACAT	1828
Oy	1261	tcaaaaacttggaaccaataaaaatacaaaaagaagaagaatagcagatcaattltgtga	1320
Db	1829	TCAAAAATCTGGACCACTAAAAATACAGAAAGACGAAAATATACAGTTCAATTTTGGA	1888
Oy	1321	aataatgtatctccgcttltcaaatgaatgtctactcttgaatagaaggggtacatc	1380
Db	1889	AATATATGATTTCTCCCACTTCAAAATGAATGTCTACTTTTGTATAGAAAGAGTACACTC	1948
Oy	1381	agtagaaatttgcaaaactctcgaagaaggtttaaagtgtaattctcgtgaaatctggccac	1440
Db	1949	AGTAAATTTGGCAAACTCTCGAACAAAGTTTAAATAGTGAATTTCCCTGTGACATCTGGCAC	2008

OY	1441	gaagaaaccctggaactctltggaactcttcagaacacattctctataacaaaggaataacacac	1500
Dp	2009	caaacactcggaacctttttggatctcttttagacacattctttatcaaaagaaacnattacattac	2068
OY	1501	tatgaatgtaaacagatgctctactctttagaagacatacaaaatctaaatctccctggtgat	1560
Dp	2069	tatgaatgcttaagcagcttcttactttgcacatcatcaaaaattctaaattctggttcacatt	2128
OY	1561	aaacacatcqlaaatctccctctctctcagagaaagacatgagatattgtacccctccaaat	1620
Dp	2129	aaacacatcgttgatttcccttctctgagcacaagatctcgtgatattcttaccctccaaatt	2188
OY	1621	agttgtatgtagacctctctaaactctgaatgaatgaaacagagatctctgaagactctacag	1680
Dp	2189	ctgtttctgttatgacgctttacacttgatggaacagagactactacaaacccattctcatgc	2248
OY	1681	gattctgaacaaatctctccagatctcagatcagacatcagaagttctgtaattgattgaactcag	1740
Dp	2249	gattctgacaaattctctccaaatcccacgactcagaaactctttaaagcatgatatgaccttcag	2300
OY	1741	aaagatagtgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat	1800
Dp	2309	aaagatagtgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat	2368
OY	1801	ccggtgctggaatgctctcatalcagaatcagaatgctcagaatggaatggaatggaatggaat	1860
Dp	2369	ccggtgctggaatgctctcatalcagaatgctcagaatggaatggaatggaatggaatggaat	2428
OY	1861	tactatcagaattctgacacacacagttcagaagaatgaatgaatgaatgaatgaatgaatgaat	1905
Dp	2429	tgcctatcagaatttttgcacacacacagtttgcacacacagtttgcacacacagtttgcacacac	2473

ORGANISM	REFERENCE	TITLE	JOURNAL	REFERENCE	AUTHORS
Human sapiens					
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:					
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.					
1 (siles)					
Matenabe, K., Kunigaki, A., Itakura, S., Yamakaki, M., Tashiro, H.,					
Ota, T., Suzuki, Y., Obayashi, M., Nishi, T., Shidabara, T., Tanaka, T.,					
Nakamura, T., Isogai, T. and Sugano, S.					
NEDO human cDNA sequencing project					
Unpublished (2000)					
2 (bases 1 to 1752)					
Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,					
Shidabara, T., Tanaka, T. and Nakamura, Y.					
Direct Submission					
Submitted (29-AUG-2000) to the DDBJ/EMBL/Genbank databases. Sumito					
Sugano, Institute of Medical Science, University of Tokyo.					
Laboratory of Genome Structure Analysis, Human Genome Center;					
Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan					
(E-mail: cdna@elms.u-tokyo.ac.jp, Tel: 81-3-5449-5286,					
Fax: 81-3-5449-5416)					
NEDO human cDNA sequencing project supported by Ministry of					
International Trade and Industry of Japan. cDNA full insert.					
sequencing. Research Association for Biotechnology: cDNA library					
construction, 5' & 3'-end one pass sequencing. Departent of					
virology and Human Genome Center, Institute of Medical Science,					
University of Tokyo (partly supported by Science and Technology					
Agency).					

FEATURES	Location/Qualifiers
source	1. 1752
	/organism="Homo sapiens"
	/db_xref="taxon:9606"

/clone="HS102573"
/clone_lib="HS1"
/tissue_type="human small intestine"
/note="Cloning vector pME185FL3"
135. 1289
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB15404.1"
/db_xref="GI:10439020"
translation="MSENQTMLEFILHSGTSYGRGIRVLPESNSDNDOLKKDLES
ANLTANQSDVICOSENPDSFSSGSLKVEKQGLSATITLDHXYLERKPLCAL
KOKAPOORIRIRAKIKRPRRLOSXKIHCKKLLQEPHEGLDITFDGATKRPD
VOLONTSLYDKSTITTKNOKRKAIVFVKNNGTILPLSNKCLLIEGCTLSEI
KFNVSITPKSCHDELIDLSAPPLIGCTTJHHCCKSSLRSTONLSLVMTSMWIP
SSVAELGIVPLOYVFWVTPTLDGCTGLEAVYLMDSKFPQIPVASEVLMDDLOKSVI
MIMDFCPGPKIDAYPMLECFIKSYNVNTMOTDQICVUIEDTTVAEDVI"

BASE COUNT 555 A 319 C 334 G 544 T
ORIGIN

Query Match 66.6%; Score 1268.4; DH 9; Length 1752;
blast Local Similarity 99.5%; Prtd. No. 8.1e-294;
Matches 1283; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

```

Oy 616 gatttaagtcacatccatcgcgtctacaaatctgacataagacatttaagctacgaataac 675
Db 1 GATTTAAGTCACATCCATCGGCTACAAA-CTGACAAATAGCAATTTTAACTTACGATNAC 59
Oy 676 cagcttcacgtgcaagaatctctgaaagcttctgaaatctctgaaatctctgaaatctct 735
Db 60 CATGTTCACTGCGCAAGATCTCTGAAAGCTTTCGAAAGCTTTCGAAAGCTTTCGAAAGCTT 119
Oy 736 accaaactcgaagaatctgaaatctgaaatctgaaatctgaaatctgaaatctgaaatctg 795
Db 120 ACCGACTTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 179
Oy 796 catggaagtaacagcttaacagcttaacagcttaacagcttaacagcttaacagcttaac 855
Db 180 CATGAGAGTACCACTTACGCTCGGGGAATCAGGCTTTCGCAAGAACTTAACCTCGATGCT 239
Oy 856 gatacaactgaaagaatctgaaatctgaaatctgaaatctgaaatctgaaatctgaaatct 915
Db 240 GATACAACTGAAAGATTTTAAATCTGCAAAATTTGACAGCAATTCAGCAATTCAGATGTT 299
Oy 916 atcgtcaatcgaacacgtgacagacgttctgaaatctgaaatctgaaatctgaaatctg 975
Db 300 ATCTGTCACTGACAGAACTGACAGACGCTTTCGAAAGCTTTCGATCAGTATCATTAATG 359
Oy 976 gtgaagaatgtaacagcttctgaaatctgaaatctgaaatctgaaatctgaaatctg 1035
Db 360 GTGAAGAATGTAACAGCTATCTGTAACAATCTTACAGATCATCTAGTATTGGAGAGG 419
Oy 1036 acaccactatgtgcatcttgaagaacaaagctcctcaacaaataacgcatccgaagaaa 1095
Db 420 ACACCACTATGTGCTATTTGAAACAAAGAGCTCTCTCAACAAATACGCAATCCAGAGAAA 479
Oy 1096 ttagagtcataaagcccgaaagactatctgaaatctgaaatctgaaatctgaaatctg 1155
Db 480 TTGAGTCATATTAAGCCCAAGACACTTTCAGCTCTTAACTTCATTCCTTAATATG 539
Oy 1156 ctttctgtgaagaagcttcaacatgagagcttgaatctgaaatctgaaatctgaaatctg 1215
Db 540 CATTGTGCAAAAGTTCCACATGAGAGGCAATTTGCAATTAATTTTTCAGAGTGTGCA 599
Oy 1216 actaaaccccaagctgtgcaagcttcaacaaatacatctatataatcaaaaactcgacc 1275
Db 600 ACTAAACCCCAATGTCCGAGCTACAAATACATCTTATATGATATCAAAATCTGAGAC 659
Oy 1276 actaaaccccaagctgtgcaagcttcaacaaatacatctatataatcaaaaactcgacc 1335
Db 660 ACTAAACCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 719
Oy 1336 cgccttcacaaatgacttacttctgaaatgaaaggtgaaacacacgagaaatctgcaaa 1395

```

```

Db 720 CCGCTTCAATGAAATGCTACTTTCATGACAGAGCTGACACTCAATTAATTTGCAAA 779
Oy 1336 ccttcgaacaaagllaaatgagtaatctcctgagagatctgagcaagaaagacgtgaact 1455
Db 780 CTCGCAACAAAGTTTAATAGCTTAATTCCTGTGAGATCTGCCACGAAAGCTGAACTT 839
Oy 1456 ttggaacttcgaacacattctctatcaagaagaacaaatacaatcaatgaaatgaaag 1515
Db 840 TTGGAACCTTTCAGACCACTTTCTTATCAAGAGCAATACATCACTATGATGATTAACG 899
Oy 1516 tctcgaacttcgaatccatcaacaaatacaatctcctgagllagataaaacatcgagat 1575
Db 900 TCTTCTGATTTGACATCTCAACAAATCTTAATTTCCCTCCTTATTAACATCTGTGAT 959
Oy 1576 ccttcctctctgagagaagacctgggtatctgaaacccctcaacaaatgagllgttctga 1635
Db 960 CTTTCTTCTGTCGACAGACAGCTGGCTATTTGACCCCTCCAAATATCTGTTTATGACC 1019
Oy 1636 tttaactttaaagtgaaacagagtaactgaagacccatctatctgaaatctgaaatctg 1695
Db 1020 TTTAACACTTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 1079
Oy 1696 ttcctgaatcttcgaacacagagatctcctgagllagataaaacatcgagat 1755
Db 1080 TTTCAGATTTCCACCAATCAGAACTTCTCATGCAATGATGATGATGATGATGATGATG 1139
Oy 1756 atcagagatctctcctcctcgaagaataaaatctgaaatctgaaatctgaaatctg 1815
Db 1140 ATCACTGATATATTTCTCTCTCCACAGATTAATAATTTGATGATGATGATGATGATG 1199
Oy 1816 tcaatcaatctcgaacacagagatctcctgagllagataaaacatcgagat 1875
Db 1200 TTCACTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 1259
Oy 1876 gacacacagcttcgaagaatgtaactcaaa 1905
Db 1260 GACACCACTTCCAGACAGATTAATCTAA 1289

```

RESULT 7
LOCUS HSM800212
DEFINITION Homo sapiens mRNA: cDNA DKFZP586D211 (from clone DKFZP586D211);
ACCESSION AL050120
VERSION AL050120.1 GI:4884145
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2012)
AUTHORS Wambut, K., Heubner, D., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-1999) MIPs, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email S. Wiemann@dkfz-heidelberg.de;
sequenced by ACOWA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKFZP586D211) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcencentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at <http://www.mips.biochem.mpg.de/proj/cDNA/>.

FEATURES

source
1. 2012
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZP586D211"
/clone_lib="586 (synonym: hute1). Vector pSport1; host
DH10B; sites NotI, SmaI/MluI"

```

/dev_stage="adult"
/tissue_type="uterus"
/map="530.1 CR from top of Chr7 linkage group"
1. .411
/genes="DKF2p586D211"
<1. .411
/genes="DKF2p586D211"
/notes="WUGSC:H_DJ0907C10.1, partly"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB43281.1"
/db_xref="GI:4884146"
/translation="DHGCKKCCSSLRSLQNSLVDTKSWIPSSVAEALGIVPLQYVF
WMFLDDGTGLAYLMDSDKFFQIPASEVLMDDLOKSDVMIMDFCPCPGIKIDAY
PWLCEFTKSYNVTNGTDNQICYQIFDTTVAEDVT"
polyA_signal
polyA_site
1965. .1970
BASE COUNT 631 a 343 c 357 g 681 t
ORIGIN

Query Match      21.4%  Score 408.4;  DB 9;  Length 2012;
Best Local Similarity 99.8%;  Pred. No. 1.5e-87;
Matches 409;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0;

QY 1495 atcaactatgatgtaaaacagtggttctggtgagatccatacaaaatctaaattccctgg 1555
|||||
Db 2 ATCAGTATCGGATGTAACAGTGTTCTAGTTTGAGATCCATACAAATCTAANTTCCCTGG 61

QY 1556 ttgataaaacatcgtggattcctctctctgtgagagagcactgggtattatccctcc 1615
|||||
Db 62 TTGATAAAACATCGTGGATCTCTCTGCTGGCAGAGCACTGGTATTCTACCCCTCC 121

QY 1616 aatatatatttattatgacctttacatttgaatgaagaaaggaafactaaagaaacctatc 1675
|||||
Db 122 AATATCTGTGTTATGACCTTTACCTTGATGATGGAACAGAGTACTAGAGCCCTATC 181

QY 1676 tcatgattctgacaaattcttcagattccagatccagagagttctatgaatgaatacc 1735
|||||
Db 182 TCATGGATCTGCAGAAATCTTCAGATTCAGACATCGAAGTTCTGATGATGATGACC 241

QY 1736 ttcaaaaaatgtagatgataatgataatgttttttctctcccaagaaatgaaattgatg 1795
|||||
Db 242 TTACAGAAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 301

QY 1796 catatccatgattgaatgattctcaatgaatgatacaaatgtacaaatgaagaaatgaatc 1855
|||||
Db 302 CATATCCGTGGTGGATGCTTCATCAAGTCAATCAATGTCACAAATGGAACAGATAATC 361

QY 1856 aaattgctatcagattttgacaccacagttgagagagtgtaattctaa 1905
|||||
Db 362 AATTGCTATCAGATTTTGCACACACAGTTGCGAGAGATGTANTCTAA 411

RESULT 8
AC004925 Locus Homo sapiens PAC clone RP5-907C10 from 7431-3412, complete
DEFINITION
AC004925 sequence.
AC004925.1 GI:4156174
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 154959)
AUTHORS Jones,K., Tin-Wollam,A. and Keppler,D.
TITLE The sequence of Homo sapiens PAC clone RP5-907C10
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 154959)
AUTHORS Waterston,R.H.
TITLE Direct Submission

```

JOURNAL

REFERENCE

AUTHORS
TITLE
JOURNAL

REFERENCE

AUTHORS
TITLE
JOURNAL

REFERENCE

AUTHORS
TITLE
JOURNAL

COMMENT

Submitted (12-JUN-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 154959)
Waterston,R.H.
Direct Submission
Submitted (14-JAN-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 154959)
Waterston,R.
Direct Submission
Submitted (22-JAN-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
5 (bases 1 to 154959)
Waterston,R.
Direct Submission
Submitted (21-DEC-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jan 14, 1999 this sequence version replaced gi:3213072.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics

Center project name: H_DJ0907C10

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NIHRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nihri.nih.gov/DIR/DTB/CHR7>, send mailto:cgreen@nihri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-5, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://haepac.mad.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.

The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong. VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of RP5-907C10; actual end is at 154959 of RP5-907C10.

The sequence from 4689 to 5920 is from PCR product from PAC RP5-907C10 DNA.

Location/Qualifiers

1. .154959
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q31-3q32"

FEATURES

source

Query Match 1.8%; Score 34.8; DB 4; Length 1046;
Best Local Similarity 50.0%; Prod. No. 3.3;
Matches 87; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
Qy 1285 caaaaagacgaaagatgaacattcatttttggaaaaataatgtatttcgcgtttca 1344


```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patcldm Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/381,862
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 1997-71077
FILING DATE: 25-MAR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JPT/JP96/07288
FILING DATE: 24-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: GAWLEY, J.F., THOMAS A.
REGISTRATION NUMBER: 40,944
REFERENCE/DOCKET NUMBER: 19036/36274
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0648
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 7143 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes
STRAIN: Clinical isolate SP-26-36
US-09-381-862-4
```

```
Query Match 1.8% Score 34.22 DB 4: Length 7143;
Best Local Similarity 50.98% P-val: No. 9, 55
Matches 81: Conservative 0: Mismatches 78: Indels 0: Gaps 0:

OY 1364 taagaagaaatgaacaaatgaatctgaacactctgaacaaatlaaataaactc 1423
DB 5125 TACAGCTGCTAAAGCTACAGCAAGTTATGATTTGATTTTTCAGCGCTGAT 5184
OY 1424 ctatgaatctgaacaaatgaacactctgaacactctgaacactctgaac 1483
DB 5185 CTGAGATATGACTTTGGAATCAACACAGACGCTGATCTTGAACTGTTTCTAC 5244
OY 1484 aaagaaacaaacaaacaaacaaacaaacaaacaaacaaacaaacaaacaa 1522
DB 5245 AGCTATATTTCTTAAGAAAGAGCTAAACATGCTTTTA 5283
```

```
RESULT 9
US-08-836-325-13
Sequence 13, Application US/08836325
Patent No. 6110672
GENERAL INFORMATION:
APPLICANT: Mandel, Gail
APPLICANT: Halegoua, Simon
APPLICANT: Horden, Laurence A.
TITLE OF INVENTION: Peripheral Nervous System Specific
TITLE OF INVENTION: Sodium Channels, DNA Encoding Theretof, Crystallization,
TITLE OF INVENTION: X-ray Diffraction, Computer Molecular Modeling, Rational
TITLE OF INVENTION: Drug Design, Drug Screening, and Methods of Making and Using
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 New York Ave., N. W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patcldm Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836,325
FILING DATE: 2-MAY-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JPT/US95/14251
FILING DATE: 02-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/482,401
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/134,029
FILING DATE: 02-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ludwig, Steven R.
REGISTRATION NUMBER: 46,203
REFERENCE/DOCKET NUMBER: 0917,0240002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-471-2600
TELEFAX: 202-471-2540
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 6371 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: DNA (genomic)
US-08-836-325-13
```

```
Query Match 1.8% Score 43.85 DB 3: Length 6371;
Best Local Similarity 53.48% P-val: No. 12;
Matches 71: Conservative 0: Mismatches 62: Indels 0: Gaps 0:

OY 379 actgaagaaacaaatgaatgaacactctgaacactctgaacactctgaac 438
DB 2320 ACTGAGGAATTCAAAGATGCTGATACCAATTTGGCTTTTACGGAATCTT 2379
OY 439 ccttgaacaaacaaacaaacaaacaaacaaacaaacaaacaaacaaacaa 498
DB 2380 GCTGAATGCTATTAAGCAATTCGCAATGCAATGCAATGCAATGCAATG 2439
OY 499 gagcctctgaagca 511
DB 2440 AATATTTTTCACA 2452
```

```
RESULT 10
US-08-836-325-14
Sequence 14, Application US/08836325
Patent No. 6110672
GENERAL INFORMATION:
APPLICANT: Mandel, Gail
APPLICANT: Halegoua, Simon
APPLICANT: Horden, Laurence A.
TITLE OF INVENTION: Peripheral Nervous System Specific
TITLE OF INVENTION: Sodium Channels, DNA Encoding Theretof, Crystallization,
TITLE OF INVENTION: X-ray Diffraction, Computer Molecular Modeling, Rational
TITLE OF INVENTION: Drug Design, Drug Screening, and Methods of Making and Using
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 New York Ave., N. W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
```


[illegible]

RESULT 2
BL256303

QY 1442 aaagccctgaaacclllagaaactllcaaacaccccllclclalacaaagaaacalacacact 1501
 DB 421 aaagacgtgaaactttttgacacctttgacacacacatttttttttttttttttttttt
 QY 1502 atgagctgaaacagatgctcagctttagaaacacaaacaaatcllcaaacllccctgllgaa 1561
 DB 481 atgagatgaaacagcttt
 QY 1562 aaagacatgaaaccllccctcllclclclclclclclclclclclclclclclclclclcl 1621
 DB 541 aaacatgctgaaacatt
 QY 1622 ttt 1681
 DB 601 ttt
 QY 1682 atcttgaacaaatcttctgaacattctgaacattctgaacattctgaacattctgaac 1740
 DB 661 attctgcaaaatt 1720
 QY 1741 aaagatgaaaccllcl 1800
 DB 721 aaagatgaaaccllcl 1780
 QY 1801 cggc-tgattgaaatgcttctgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaat 1857
 DB 781 cccgtgctgaaatgcttctgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaat 1838
 RESULT 3
 BE891252 926 bp mRNA EST 20-OCT-2000
 LOCUS 601432077F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3917299 5'
 DEFINITION mRNA sequence.
 ACCESSION BE891252
 VERSION BE891252.1 GI:10450795
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 NIH-MGC http://mrc.nhl.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaups-f@mail.nih.gov
 Tissue Procurement: ATCC/ICM/DTP
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The J.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Inyece Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the J.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: L1AM9743 row: 3 column: 20
 High quality sequence stop: 636.
 location/Qualifiers
 FEATURES
 SOURCE
 1. 926
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3917299"
 /clone_1="NIH_MGC_72"
 /issue_type="melanotic melanoma"
 /lab_host="p310b (phage-resistant)"
 /note="Organ: skin; Vector: pCMV-SORTO; Site_1: NotJ;
 Site_2: SalI; Cloned unidirectionally. Primer: OJuo d1.
 Average insert size 2 kb. Library constructed by Life
 Technologies."
 BASE COUNT 273 a 205 c 212 g 238 t
 ORIGIN
 Query Match 32.1%; Score 611.2; DI 11; Length 926;

Best local similarity 94.0%; Pred. No. 8, 1e-137;
 Matches 679; Conservative 0; Mismatches 38; Indels 5; Gaps 4;
 QY 87 aaagctclttaaaccacccatcattcagaaagaaacagaaacagaaacagaaacagaaacagaaac 146
 DB 1 GAAGTCTTTAAGCCCCCATATCTAAGCAAGAAAGAACTGATTTATCTCAGTTGTAACATAT 60
 QY 147 ttt 206
 DB 61 ttt 120
 QY 207 ttt 266
 DB 121 ttt 180
 QY 267 ttt 326
 DB 181 ttt 240
 QY 327 ttt 386
 DB 241 ttt 400
 QY 387 ttt 446
 DB 401 ttt 460
 QY 447 ttt 506
 DB 361 ttt 420
 QY 507 ttt 566
 DB 421 ttt 480
 QY 567 ttt 626
 DB 481 ttt 540
 QY 627 ttt 686
 DB 541 ttt 596
 QY 687 ttt 745
 DB 600 ttt 706
 QY 746 ttt 805
 DB 657 ttt 716
 QY 806 ttt 866
 DB 717 ttt 826
 RESULT 4
 B1257710 1159 bp mRNA EST 17-JUL-2001
 LOCUS 602967368F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5106881 5'
 DEFINITION mRNA sequence.
 ACCESSION B1257710
 VERSION B1257710.1 GI:14813351
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 1159)
 AUTHORS NIH-MGC http://mrc.nhl.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.

Email: c9apbs+email.nih.gov

Tissue Procurement: ARCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: Incyte Genomics, Inc.

cDNA Sequencing by: Incyte Genomics, Inc.

clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

plate: FLAM11258 row: h column: 18

High quality sequence stop: 705.

Location/Qualifiers

1. il159

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5106881"

/clone_lib="NIH_MGC_12"

/tissue_type="cervical carcinoma cell line"

/lab_host="DH10B"

/note="Organ: cervix; Vector: pCMV-Sport6; Site:1: NotI; Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.4 kb. Library prepared by Life Technologies."

Query Match	30.8%	Score 587	DB 11	Length 1159	
Host Local Similarity	91.1%	Pred. No. 5,60-131			
Matches 706	Conservative	0	Mismatches 50	Indels 19	Gaps 7
766	cgagcaatgtaaatcttlaagaatttctctctctatcaatgaattaccattacagttccggaatc	825			
2	CAGACAATGTTAACTTTTAGAGATTTTCATCTCATGGAGGTACCAGTTACGGTCCGGGAATC	61			
826	agggcttgcagaaataactctgattggatcaactgaataaagaatttctgaatctaca	885			
62	AGGTTCTTCCAGAAAGTAATCTGATGTGGATCAACTGAAAAGAGATTATGAATCAGCA	121			
886	aattgcacagcaatcagcatcagatgattatctctcaatcagaacctcagacagctctt	945			
122	AATTTGACAGGCAATCAGCAATCAGATGTTATCTGTCAATCAGAACCTCAGCACACGCTTT	181			
946	ccaaactctggatcagtatctattatcaagagtaagaagaatgtaaacagctatctctaca	1005			
182	CCAAGCTCTGGATCAGTATCATTAACGAGGTAGAAAAGATGTCACAGAGCTATCTGCTACA	241			
1006	atacttcaagatcatcagattcttggagagagagaccactatgagcatctctctacaacaaa	1065			
242	ATACACTTACAGATCATCAGTATTTGGAGAGGACACCACTATGTGCCCATTTTGAACACAAA	301			
1066	gctctctcaacatacaccgatactcgaagcaaatctgaattcataatggcccaagagctctt	1125			
302	GCTCTCAACAAATACCGCATCCGAGCAAAATTCAGATTCATATTAAGCCCAAGACATATTT	361			
1126	caatctgttaaaactctcatctgctcaatgatgcaattctctcaagaaagtctctacataaggg	1184			
362	CAGTCTGTTTAAACTTTCATTTGCCCTTAAATGTGATTTTTCGTCACAAAGATTTCCACATGACGG	421			
1185	gaaatttgatataatcttctgaatgattgcaactaaatcccaattctctcagttacaaa	1243			
422	CGATTTTGGATATAATTTTTCAGATGCTGCACATAAAAGCCAGATGTCAGACTACAAA	481			
1244	atacactcatatatagatataaaaatctcgagaccactaaaatc-aaaagaagcaaaaagta	1302			
482	ATACATCATATATGATTTCAAAAATCTGGAGCACATAAATTCACAAACGACGAAAGTA	541			
1303	gagtttcatctttctg-aaaaaataatggatctctcccgctttccaatgaatgctctactt	1360			
542	CGAGTTCACTTCTGCACACAATTAATGGTTATCTCCCGCTTCCAAATGAATGTCTACTTT	601			
1361	tgataagaaggaggtacactcagtaaaa-ttgcgaactctcgaacaagttaaataatgta	1419			
602	TCATAGAAGGAGGTACACTCAGTGAATTTTGCAACTCTCGAACAGTCTAATAGTGTGA	661			

1420	attcctatgaga-----tctggcaccgaagaccttggagcttttcagcac	1471
Qiy		
1430	tt	
Ddb		
662	anttcctctcagcagatcttgggcacccgaagcaaaccttggcagacttttggacac	721
Qiy		
1472	ca-----tctctatatacagaacaatacatcatcactatgatatgaacagtctt	1521
Ddb		
722	caatttctttttatgtccagcgagccatcacatcaattctacatgatgtctcacacagtgct	776
Qiy		
LOCUS	BE884820	
DEFINITION	601510547f1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3912010 5',	20-OCT-2000
ACCESSION	BE884820	EST
VERSION	BE884820.1	mrna
KEYWORDS	BE884820.1 GI:10333596	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE	1 (bases 1 to 588)	
JOURNAL	NIH-MGC http://mgc.ncl.nih.gov/ .	
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:	

FEATURES
SOURCE

BASE COUNT
ORIGIN

Query Match	30.6%	Score 582.8;	DB 11;	Length 588;
Best. Local Similarity	99.7%	Prod. No. 5,40-130;		
		0 mismatches	2 indels	0 gaps

[illegible]

DB
182 TTACAGATCATCCTATTTCGACAGGACACCACTATGTGGCATTTTGGAGCAAAATTT

```

CY 1070 ctcacacataccgcacccgagcgaaballagagctcatalaagccgcgaagacatallacgaat 1129
DB 242 CTCACAAATACCCGACCCGACCAAAATTCAGCTATTAAGCCGACACACATATTTTCACAT 301
CY 1130 cctctaaactccttaccctaaatgctcattctcgcgaagaagctccacatagagcgaatt 1189
DB 302 CTCCTTAATCTTCATTCGCCCTAAATGTCATTTCTGCGACACAACTTCGACATGACGCCGAT 461
CY 1190 agagatcaatcttcagagatggtcgaactaaagcccccattatcgaagcttgcgaatcaat 1249
DB 362 TCGATATTAATTTTTCAGATGCGTCGAACATAAAGCCGACATTCACACCTCAAAATACAT 421
CY 1250 cctcatalaattcaaaaactctgagccgactaaataatcgaagaagcgaagaagatgagcttc 1309
DB 422 CATTAATATCTATTAATAAATCTGACATACATAAATATCAAAAGACACAACTGACATTC 481
CY 1310 attctatgaagaaataatgataatctctccgcctctcgaatgaattctgcttgcgaagaa 1369
DB 482 ATTTTGTGAAAATATATGATGATTTCTCCGCTTTCAAAATGATCTCTATCTTTGATATGAAG 541
CY 1370 gaagtagactcagtguaallgcgaactctcgaagaatttcaaaag 1415
DB 542 GAGGTACACTTCACATCAATTTCTCAAACTTCCTGCAACACAACTTAAATAC 597
RESULT 6
LOCUS HG390611 584 bp mRNA EST 12-MAR-2001
DEFINITION 6021639131 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4524671 c.
ACCESSION BC390611
VERSION BC390611.1 GI:13284059
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 584)
AUTHORS NIH-MGC http://mgi.sci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rstraub@mail.nih.gov
Tissue procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The J.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the J.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: L1AM10429 row: c column: 24
High quality sequence stop: 582.
Location/Qualifiers
source
1..584
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4524671"
/tissue_type="Embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-SPORT6; Site: 1; Note:
Site 2: Salt: Cloned unidirectionally; clone-ct primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: This is a NIH-MGC library."
BASE COUNT 198 a 114 c 105 g 163 t
ORIGIN

```

Query Match

Best Local Similarity 99.5%; Freq. No. 40-129; Length 584;
Matches 581; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

CY 780 ttttgaagcttctcctcctcctgaagagctacagctcagaggaatcgaagcttctacagga 839
DB 1 TTTTACATCTTTCATCTCTTCAGGAGACCACTTACGCTCCGCGAATCAGCTTCTGCGACA 60
CY 840 aagtaactctaatatgaatgaactgaagaagactllagaatctgcgaacttgatgaagcga 899
DB 61 AAGTAACCTCTCATCTGCGATGATCAAGTAAAGGATTTGCAATTCGCAATTTGACACGCA 120
CY 900 tgcgaactgcgaatgctcctgaatgcgaatgcgaagcctgcgaagccttgcgaagctgcgaatc 959
DB 121 TCGACATTCACATCTTATCTCTCAATGACAAATCTGCACGACGCTTCCAAAGCTCTGATAC 780
CY 960 gattatcattatgcgaagatgaagaatgattgcgaagcattgctgaactgctgcgaatgctc 1019
DB 181 ACTATATATATACGACATACAAAGATCTGACAGCTATCTGCTATACATATTAACAGATGA 240
CY 1020 tgcgaatcttgaagaagacagcgaatgattgctgcttgaagaagacgctgcgaagaa 1079
DB 241 TCGATATTTTCAGACGACACGACGATGCTGCTCATTTTCGCAAAAGCTCTGACACATA 400
CY 1080 cgcgaatgcgaagcgaatctgaagcattatgcgaagcgaagcatttgcgaatctgaact 1139
DB 401 GTCATATCTCACTAAATAATTCACCTGATATTAAGCTGCACAACTATTTTCATCTTAAAT 360
CY 1140 tcatgctcctgaatgctgcttgcgaagaagctgcgaatgaagcgaattggaatgatt 1199
DB 461 TCAATTCCTCTAATATCTATTTCTGCAAGAACTTCGACATGACAGCTATTTGCAATTAAT 420
CY 1200 ttttgcgaatgattgcgaatgaagctgctgctgaattgcgaatgcgaatgctcattatga 1259
DB 421 TTTTTCAGATCTGCTGCACTTAAATCTCCGATGTCGACCTTCGCAAAATGATCATATATCA 480
CY 1260 ttcgaagaatctggaagcctcgaagaatcgaagaagcgaagaagtagcaatllctatga 1319
DB 481 TTTCAAAATATCTTCGACCTCAAAATTCGCAAGCAAGCAAGCTTCGACCTTATTTCTTAA 540
CY 1370 aagtaactgattctcgcgcctcgaagaaatctgcgaatttgaag 1463
DB 541 AATATATGCTTATCTCCGCTTCAATGATGCTACTTTTGA 584
RESULT 7
LOCUS BE295350 736 bp mRNA EST 20-JUL-2000
DEFINITION 60117585431 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:3531496.5.
ACCESSION BE295350
VERSION BE295350.1 GI:9178898
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 736)
AUTHORS NIH-MGC http://mgi.sci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rstraub@mail.nih.gov
Tissue procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The J.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the J.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: L1CM205 row: k column: 13
High quality sequence stop: 655.
Location/Qualifiers
source
1..736
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3531496"
/tissue_type="Embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-SPORT6; Site: 1; Note:
Site 2: Salt: Cloned unidirectionally; clone-ct primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: This is a NIH-MGC library."
BASE COUNT 198 a 114 c 105 g 163 t
ORIGIN

```

(HXSE-) HXSEO INC.

PA Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D:
PI Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J:
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT:

XX WPI: 2001-442253/47.
DR N-PSDB: AA159281.

PI Novel nucleic acids and polypeptides, useful for treating disorders
PI such as central nervous system injuries -

XX Example 5: SEQ ID NO 3270: 10078pp; English.

XX The invention relates to human nucleic acids (AA157798-AA15799) and
CC the encoded polypeptides (AA158542-AA158543) with neurotrophic
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX Sequence 524 AA:

Query Match 96.5%; Score 2637; DB 22; Length 634;
Best Local Similarity 99.2%; Pred. No. 9, 5e-263;
Matches 503; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

DB 1 MSLVFAINVIYPLNOLKGGTIVNVGVKFFKPPYLSKGTIDYCSVTVIVDOTNKLITCL 50
OY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 1 MSLVFAINVIYPLNOLKGGTIVNVGVKFFKPPYLSKGTIDYCSVTVIVDOTNKLITCL 50
OY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 61 LFSGNVEALPIYKNGDIYFRHLKIQVKKETOGIISGFAISLFEGLTGAPLIPRIS 120
OY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 61 LFSGNVEALPIYKNGDIYFRHLKIQVKKETOGIISGFAISLFEGLTGAPLIPRIS 120
OY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 121 KTFNFTEDHKVYEAIRVWASTHMSKWLKLDVQPMOYEDLTQILGKAEVQGSFL 180
OY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 121 KTFNFTEDHKVYEAIRVWASTHMSKWLKLDVQPMOYEDLTQILGKAEVQGSFL 180
OY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 181 LKVDGTFPFPSPWYLIDVLEGDLSHHPLONLIDILVYDHHVHARSILKVGSLP 240
OY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 181 LKVDGTFPFPSPWYLIDVLEGDLSHHPLONLIDILVYDHHVHARSILKVGSLP 240
OY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 241 IYSLHTIKLOSMSNENITMSLEFHLHGCTISYGCIRVLPESNDVQLKDLFSAULTAH 300
OY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 241 IYSLHTIKLOSMSNENITMSLEFHLHGCTISYGCIRVLPESNDVQLKDLFSAULTAH 300
OY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 301 QHSDVTCGSEPPDSFSSGVSLSYEVRCOOLATILTDHQTLEPRLPACAILKOKAPQOY 360
OY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 301 QHSDVTCGSEPPDSFSSGVSLSYEVRCOOLATILTDHQTLEPRLPACAILKOKAPQOY 360
OY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 361 RIRAKLRSTYKPRRLFQSVKLHCPKCHLLOEVRPHEGDLIDFDDGATKTPDVKLONTSLYD 420
OY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 361 RIRAKLRSTYKPRRLFQSVKLHCPKCHLLOEVRPHEGDLIDFDDGATKTPDVKLONTSLYD 420
OY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 421 SKITWITKNGKRVAAHVEFVNNGLPLSRCLLIEGGLTSEICKLSHRSYFVRSQH 480
OY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 421 SKITWITKNGKRVAAHVEFVNNGLPLSRCLLIEGGLTSEICKLSHRSYFVRSQH 480
OY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 481 EDLELDLSAPFLIOGTIHHYGTGYCT 507
OY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 481 EDLELDLSAPFLIOGTIHHYGTGYCT 507
OY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 2

AA093478
ID AA093478 standard; Protein: 634 AA.

XX AA093478:

AC 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:12761.

XX Human: primer: detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000: 2000EP-0116126.

XX 29-JUL-1999: 99JP-0248036.

XX 27-AUG-1999: 99JP-0300253.

XX 11-JAN-2000: 2000JP-0118776.

XX 02-MAY-2000: 2000JP-0183767.

XX 09-JUN-2000: 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J:

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T:

XX WPI: 2001-318749/34.

XX Claim 8: SEQ ID 12761: 2537P + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification, where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences: AA092446 to
CC AA095893 represent human amino acid sequences: and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 534 AA:

Query Match 96.5%; Score 2637; DB 22; Length 634;
Best Local Similarity 99.2%; Pred. No. 9, 5e-263;
Matches 503; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
OY 1 MSLVFAINVIYPLNOLKGGTIVNVGVKFFKPPYLSKGTIDYCSVTVIVDOTNKLITCL 60

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 17:08:13 ; Search time 58.75 seconds
(without alignments)
653,106 Million cell updates/sec

Title: US-09-816-248-17
Perfect score: 2732
Sequence: 1 MSIVPATNRYITPLNOLKGG.....HHYGTGCTPICVCYDLYT 518

Scoring table: BLOSUM62
Gapop 10.0, Capext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A-Geneseq-1101.*
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2637	96.5	634	22	Human polypeptide
2	2637	96.5	634	22	Human protein sequ
3	2637	96.5	634	22	Human protein sequ
4	2628	95.2	634	22	Human protein sequ
5	922	33.7	307	22	Human polypeptide
6	155	5.7	30	22	peptide -8901 enco
7	125	4.6	467	21	Arachidonic thal
8	105	3.9	2391	15	Cardiac 1-phosphat
9	105	3.8	555	15	Serine kinase SRP
10	104	3.8	1112	21	Haemo: nagic enter
11	102	3.7	4393	22	Amino acid sequenc

ALIGNMENTS

12	99	3.5	684	20	AAV34460	Porphyromonas ging
13	99	3.6	686	20	AAV34336	Porphyromonas ging
14	99	3.6	1002	21	AA38741	Arabidopsis thalia
15	98.5	3.5	653	22	AA847250	Human p7. Homo s
16	98	3.5	872	22	AAU07149	C. elegans CRIM1 P
17	97	3.5	372	19	AAW98360	H. pylori GHP. 134
18	97	3.5	372	22	AAW60318	Helicobacter pylor
19	97	3.6	1123	21	AA35865	Arabidopsis thalia
20	96	3.5	1139	21	AA318793	Amino acid sequenc
21	95.5	3.5	332	21	AA359480	Arabidopsis thalia
22	95.5	3.5	354	21	AA359488	Arabidopsis thalia
23	95.5	3.5	357	21	AA359488	Arabidopsis thalia
24	94.5	3.5	1568	20	AAV13462	Viral encoded sema
25	94.5	3.5	1568	22	AAW28522	Human VESPR. Homo
26	94.5	3.5	1568	22	AAU00239	Viral encoded sema
27	94.5	3.5	1568	22	AAU00435	Human VESPR. Homo
28	94.5	3.5	1568	22	AAW70131	Amino acid sequenc
29	94.5	3.5	1568	22	AAW31693	Human VESPR. Homo
30	94	3.4	467	21	AAW23071	Viral encoded sema
31	94	3.4	536	21	AAW31373	Human VESPR. Homo
32	94	3.4	537	21	AAW23070	Arabidopsis thalia
33	94	3.4	557	21	AAW23069	Arabidopsis thalia
34	94	3.4	606	21	AAW31372	Arabidopsis thalia
35	94	3.4	626	21	AAW31371	Arabidopsis thalia
36	94	3.4	1139	21	AAW18794	Amino acid sequenc
37	94	3.4	2285	20	AAW98149	Bacillus subtilis
38	93.5	3.4	519	21	AAW41694	Human secreted pro
39	93.5	3.4	1461	19	AAW64468	Human C420.2 pro
40	93.5	3.4	1461	22	AAW90743	Human PRO326 prote
41	93	3.4	1119	20	AAW08114	Human PRO326 poly
42	93	3.4	1119	21	AAW70674	Human PRO326 poly
43	93	3.4	1119	22	AAW12347	Human immune resp
44	93	3.4	1119	22	AAU00827	Human immune resp
45	93	3.4	1119	22	AAW80263	Human PRO325 prote

ALIGNMENTS

RESULT 1

AAW40125

AAW40125 standard: Protein: 634 AA.

AAW40125:

22-OCT-2001 (first entry)

Human polypeptide SEQ ID NO 3270.

Human: nocotropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathic; Huntington's disease; haemostatic; Alzheimer's; Parkinson's disease; Huntington's disease; Chemotactic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; Chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukaemia.

Homo sapiens.

W0200153712-A1.

26-JUL-2001.

26-DEC-2000: 2000WO-US34253.

21-JAN-2000: 2000US-0488725.

23-APR-2000: 2000US-0552317.

PR 09-JUL-2000: 2000US-0598042.

FR 19-JUL-2000: 2000US-0620312.

FR 03-AUG-2000: 2000US-0653450.

14-SEP-2000: 2000US-0662191.

PR 15-OCT-2000: 2000US-0693036.

FR 26-NOV-2000: 2000US-0727344.

RESULT 1
AAW40125 standard: Protein: 634 AA.

AAW40125: (first entry)

22-OCT-2001 (first entry)

Human polypeptide SEQ ID NO 3270.

Human: nocotropic; immunosuppressant; cytostatic; gene therapy; cancer;

peripheral nervous system; neuropathy; central nervous system; CNS;

Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

amyotrophic lateral sclerosis; Shy-Drager Syndrome; Chemotactic;

chemokine; thrombolytic; drug screening; arthritis; inflammation;

leukaemia.

Homo sapiens.

W0200153312-A1.

26-JUL-2001.

26-DEC-2000: 2000WO-US34263.

21-JAN-2000: 2000US-0488725.

25-APR-2000: 2000US-0552317.

09-JUL-2000: 2000US-0598042.

19-JUL-2000: 2000US-0620312.

03-AUG-2000: 2000US-0553450.

14-SEP-2000: 2000US-0662191.

15-OCT-2000: 2000US-0693036.

25-NOV-2000: 2000US-0727344.

us-09-816-248-17.rsp

[illegible]

```

RESULT      14          PRI:   1044 AA.
O9Y4E4      PRELIMINARY:
O9Y4E4      O9Y4E4      Created)
ID           O9Y4E4: 1999 (ITEMBLrel. 12, last sequence update)
DI           01-NOV-1999 (ITEMBLrel. 17, last annotation update)
DT           01-JUN-2001 (ITEMBLrel. 17, last annotation update)
DE           KIA00560. PROTEIN (FRAGMENT).
GN           KIA00560.
GE           Homo sapiens (Human).
OS           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC           Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX           NCBI_TaxId=9606;
RN           SEQUENCE FROM N.A.
RP           TISSUE=BRAN; PubMed=6628561; Tanaka A., Kotani H., Nomura N.
RC           MEDLINE=9820545; Miyajima N., Nagase I., Ishikawa K.,
PX           Nagase I., Ishikawa K., Miyajima N., Nagase I., Ishikawa K.,
RA           Ohara O.;
RI           The complete sequences of the coding sequences of 100 new cDNA clones from human which can
RT           code for large proteins in vitro." ;
RL           DNA Res 5:31-39(1998).
DR           EMBL: AB011152; BA025506.1;
DR           InterPro: IPRO00196; RhOGAP.
DR           InterPro: IPRO01154; Znf-GCS.
DR           InterPro: IPRO01849; PH.
DR           InterPro: IPRO0159; PH. 4.
DR           Pfam: PF00169; RhOGAP. 1.
DR           Pfam: PF00620; RhOGAP. 1.
DR           Pfam: PF01412; ArGININACTING.
DR           Pfam: PR00405; REV-DOMAIN. 3.
DR           PRINTS: PSS0003; PH_DOMAIN. 1.
DR           PROSITE: PSS0003; ArGAP. 1.
DR           SMART: SM00233; PH. 3.
DR           SMART: SM00324; RhOGAP. 1.
DR           SMART: SM00324; RhOGAP. 1.
FT           NON_TER
SO           SEQUENCE      1044 AA: 119463 MW: 68E2ZA6EDAA53BD CPO64;

Query Match      4.0%      Score 110.5; DB 4; length 1044;
Best Local Similarity      20.2%; Pred. No. 1.7;

```

Query Match	4.0%	Score 10
Best Local	30.8%	
Matches	110	
Similarity		

Conservative 86; Predicted No. of DB 5; Length 1217;
Mismatches 166; Indels 172; Gaps 30.

Db Oy 45 SVYIVDQIVKTLCL-----ISG-----NYEELPIYKMGDIYFHEHLYKIOYKK 91
Db Oy 483 AVISVLDOLQVHNSNLSIVKRLIRKIKQIVNWSVETLSIISAKIWS-----ELYOVOK 538
Db Oy 92 ETQIGISSGFASLTGIGIAPIRIS-----SKYFNFTIEDHKNVEALEVWASTH 143
Db Oy 535 TAENITVS-----DGNILDDLNSLVEIETSKCFEQYKTNPTLESTYSKLR-----SLR 588
Db Oy 144 MSPKTIILKICDVOPMQYFDLTCLQKGEVDCASFLKVMQST-----RIP 190
Db Oy 589 MFDGETISTM-----LEYLD-----LKGQVSNLLVEKEIISKSTSHVNSINPV-----LKP 640
Db Oy 191 FSKRYVLIIDL-----VLEGDSHT--HR-----LQNLITD-----LYVDHNVH-----228
Db Oy 641-----GALKRIGISTRYLE-DLENVROHQLFISIRNFITADVKNYGENLLEMIWPERK 655
Db Oy 229-----VARSKAGVS-----FLRYSLHTRKQSNMSNOTMLSEFHL--WGT- 269
Db Oy 656 LEELLEADOLNGFQSGSLPNSLTEMARVFEDEAKIKQIGDPRKMSANNSINILQAGTID 755
Db Oy 270-----SYRGJRYLPESHSDVDOLKQDESANIL-----ANQHSQYI--CQSEFES 314
Db Oy 756 OLASVFRFRITPELNLDFAKHEASLSKASVYISLAYVDFLFGNSKSKILLIENKIRHS 615
Db Oy 315 FF-----SSGSV-----FPMIYILLICIGSHILIGLCFCVAVAGFTKNGRAKATILLYLYFGKRLDPERK 337
Db Oy 816 FPMIYILLICIGSHILIGLCFCVAVAGFTKNGRAKATILLYLYFGKRLDPERK 337
Db Oy 338 TDHGYLERIPCA-----ILKKAPDOOVRIPAKLSYRPFIFQSYKWLCKPCHLDE-----296
Db Oy 874 LDNODGNKILLDAVPEIKNTYKWKAKVGLSYITAYNK-----EGNTHMAATKSLPEIYKI 230
Db Oy 391--VPHEDDLDIFODADATYTPDYKIONTSLYDSKIKITKIKMGGRKAVAVHNTYK-----296
Db Oy 521 LEHMGADITLL--NTLTKTIPQWITD-----

Search completed: April 3, 2002, 17:10:45
Job time: 245 sec

PR 11-MAY-1999: 99US-0134256.
PR 14-MAY-1999: 99US-0134218.
PR 14-MAY-1999: 99US-0134219.
PR 14-MAY-1999: 99US-0134221.
PR 14-MAY-1999: 99US-0134370.
PR 18-MAY-1999: 99US-0134768.
PR 18-MAY-1999: 99US-0134941.
PR 20-MAY-1999: 99US-0135124.
PR 21-MAY-1999: 99US-0135353.
PR 24-MAY-1999: 99US-0135629.
PR 25-MAY-1999: 99US-0136021.
PR 27-MAY-1999: 99US-0136392.
PR 28-MAY-1999: 99US-0136782.
PR 01-JUN-1999: 99US-0137222.
PR 03-JUN-1999: 99US-0137528.
PR 04-JUN-1999: 99US-0137529.
PR 07-JUN-1999: 99US-0137724.
PR 08-JUN-1999: 99US-0138094.
PR 10-JUN-1999: 99US-0138540.
PR 10-JUN-1999: 99US-0138847.
PR 14-JUN-1999: 99US-0139119.
PR 16-JUN-1999: 99US-0139452.
PR 16-JUN-1999: 99US-0139453.
PR 17-JUN-1999: 99US-0139452.
PR 17-JUN-1999: 99US-0139452.
PR 18-JUN-1999: 99US-0139454.
PR 18-JUN-1999: 99US-0139455.
PR 18-JUN-1999: 99US-0139456.
PR 18-JUN-1999: 99US-0139457.
PR 18-JUN-1999: 99US-0139458.
PR 18-JUN-1999: 99US-0139459.
PR 18-JUN-1999: 99US-0139460.
PR 18-JUN-1999: 99US-0139461.
PR 18-JUN-1999: 99US-0139462.
PR 18-JUN-1999: 99US-0139463.
PR 18-JUN-1999: 99US-0139750.
PR 18-JUN-1999: 99US-0139753.
PR 21-JUN-1999: 99US-0139817.
PR 22-JUN-1999: 99US-0139899.
PR 23-JUN-1999: 99US-0140353.
PR 23-JUN-1999: 99US-0140354.
PR 24-JUN-1999: 99US-0140695.
PR 28-JUN-1999: 99US-0140823.
PR 29-JUN-1999: 99US-0140991.
PR 30-JUN-1999: 99US-0141287.
PR 01-JUL-1999: 99US-0141842.
PR 01-JUL-1999: 99US-0142154.
PR 02-JUL-1999: 99US-0142055.
PR 06-JUL-1999: 99US-0142830.
PR 08-JUL-1999: 99US-0142803.
PR 09-JUL-1999: 99US-0142920.
PR 12-JUL-1999: 99US-0143542.
PR 13-JUL-1999: 99US-0143524.
PR 14-JUL-1999: 99US-0144005.
PR 15-JUL-1999: 99US-0144085.
PR 16-JUL-1999: 99US-0144086.
PR 16-JUL-1999: 99US-0144325.
PR 19-JUL-1999: 99US-0144331.
PR 19-JUL-1999: 99US-0144332.
PR 19-JUL-1999: 99US-0144333.
PR 19-JUL-1999: 99US-0144334.
PR 19-JUL-1999: 99US-0144335.
PR 20-JUL-1999: 99US-0144352.
PR 20-JUL-1999: 99US-0144632.
PR 20-JUL-1999: 99US-0144884.
PR 20-JUL-1999: 99US-0144814.
PR 21-JUL-1999: 99US-0145086.
PR 21-JUL-1999: 99US-0145088.
PR 22-JUL-1999: 99US-0145085.
PR 22-JUL-1999: 99US-0145087.
PR 22-JUL-1999: 99US-0145099.
PR 22-JUL-1999: 99US-0145192.
PR 23-JUL-1999: 99US-0145145.

PR 23-JUL-1999: 99US-0145218.
PR 23-JUL-1999: 99US-0145224.
PR 26-JUL-1999: 99US-0145276.
PR 27-JUL-1999: 99US-01452913.
PR 27-JUL-1999: 99US-0145918.
PR 27-JUL-1999: 99US-0145919.
PR 28-JUL-1999: 99US-0145951.
PR 02-AUG-1999: 99US-0146386.
PR 02-AUG-1999: 99US-0146388.
PR 02-AUG-1999: 99US-0146389.
PR 03-AUG-1999: 99US-0147038.
PR 04-AUG-1999: 99US-0147204.
PR 04-AUG-1999: 99US-0147302.
PR 05-AUG-1999: 99US-0147260.
PR 05-AUG-1999: 99US-0147302.
PR 06-AUG-1999: 99US-0147416.
PR 09-AUG-1999: 99US-0147493.
PR 09-AUG-1999: 99US-0147935.
PR 10-AUG-1999: 99US-0148171.
PR 12-AUG-1999: 99US-0148341.
PR 13-AUG-1999: 99US-0148655.
PR 13-AUG-1999: 99US-0148684.
PR 13-AUG-1999: 99US-0149368.
PR 16-AUG-1999: 99US-0149175.
PR 17-AUG-1999: 99US-0149426.
PR 18-AUG-1999: 99US-0149722.
PR 20-AUG-1999: 99US-0149723.
PR 20-AUG-1999: 99US-0149929.
PR 20-AUG-1999: 99US-0149902.
PR 23-AUG-1999: 99US-0149930.
PR 23-AUG-1999: 99US-0150566.
PR 25-AUG-1999: 99US-0150684.
PR 26-AUG-1999: 99US-0151065.
PR 27-AUG-1999: 99US-0151065.
PR 27-AUG-1999: 99US-0151080.
PR 27-AUG-1999: 99US-0151303.
PR 30-AUG-1999: 99US-0151438.
PR 31-AUG-1999: 99US-0151930.
PR 01-SEP-1999: 99US-0152363.
PR 07-SEP-1999: 99US-0153070.
PR 10-SEP-1999: 99US-0153758.
PR 13-SEP-1999: 99US-0154018.
PR 15-SEP-1999: 99US-0154039.
PR 16-SEP-1999: 99US-0154778.
PR 20-SEP-1999: 99US-0155139.
PR 22-SEP-1999: 99US-0155486.
PR 23-SEP-1999: 99US-0155659.
PR 24-SEP-1999: 99US-0156458.
PR 28-SEP-1999: 99US-0156596.
PR 29-SEP-1999: 99US-0157117.
PR 04-OCT-1999: 99US-0157753.
PR 05-OCT-1999: 99US-0157865.
PR 06-OCT-1999: 99US-0158029.
PR 07-OCT-1999: 99US-0158232.
PR 08-OCT-1999: 99US-0158369.
PR 12-OCT-1999: 99US-0159293.
PR 13-OCT-1999: 99US-0159294.
PR 13-OCT-1999: 99US-0159295.
PR 13-OCT-1999: 99US-0159329.
PR 14-OCT-1999: 99US-0159330.
PR 14-OCT-1999: 99US-0159331.
PR 14-OCT-1999: 99US-0159637.
PR 14-OCT-1999: 99US-0159638.
PR 14-OCT-1999: 99US-0159584.
PR 18-OCT-1999: 99US-0160741.
PR 21-OCT-1999: 99US-0160767.
PR 21-OCT-1999: 99US-0160768.
PR 21-OCT-1999: 99US-0160770.
PR 21-OCT-1999: 99US-0160814.
PR 21-OCT-1999: 99US-0160815.
PR 22-OCT-1999: 99US-0160980.

```

PR 22-OCT-1999: 990S-0160981.
PR 22-OCT-1999: 990S-0160989.
PR 25-OCT-1999: 990S-0161404.
PR 25-OCT-1999: 990S-0161405.
PR 25-OCT-1999: 990S-0161406.
PR 26-OCT-1999: 990S-0161359.
PR 26-OCT-1999: 990S-0161360.
PR 26-OCT-1999: 990S-0161361.
PR 28-OCT-1999: 990S-0161920.
PR 28-OCT-1999: 990S-0161992.
PR 28-OCT-1999: 990S-0161993.
PR 29-OCT-1999: 990S-0162142.

```

Query Match 4.6% Score 125 DB 21 Length 457
 Best Local Similarity 24.0% Pred. No. 0.0009
 Matches 65: Conservative 38: Mismatches 92: Indels 76: Gaps 13:

```

OY 23 VAVVGVVFFKRPPLSKGTDVCSVTIVDQT--NKLITCLLSCGNLEALPIIKKSDIVR 80
DB 24 VSLIGVLEqREPqCRMDVCLIRLIDDTYPSPqLTVVNSKLEqLPqIKNDMDML 83
OY 61 FHLRLIQVYK--KETOGITS---SGFASLFEGLGAPILPRISKEYFTIEDHK---- 131
DB 64 FRLIMqTDSqKRVNAASRVSSFA--lfeg-----vdfcygcstnfneeeal 132
OY 132 ---WREALPVKA-----STHNSP-SWILLKLDVQPMQYFDLTQQLGKALV 174
DB 133 YKsAMdLrkVfagcsqVikAMqSIeyrkpcseVsfllrElkIqgrfdlvcrlhAded 192
OY 175 DGASFLKVMQDTRTP-----FPSMRV---LIQDLVLE---GDLSIHHL 213
DB 193 tseVt--vwdgtDappasIlakrseedkafasIsVhllIsrdvllsftvgtlrlvhts 249
OY 214 QMLITIDLVYDNHVNARSLKVGSLRIYSL 244
DB 250 s-----hlfyrakpgdwklyhl 267

```

RESULT 8

AAR55694 ID AAR55694 standard: Protein: 2391 AA.

AC AAR55694:

DI 06-DEC-1994 (first entry)

XX Carbamoyl-phosphate-synthetase II.

DE Carbamoyl-phosphate-synthetase II. CPSII: pscPSII gene;

KM Carbamoyl-phosphate-synthetase II. CPSII: pscPSII gene;

XX Plasmodium falciparum.

OS Plasmodium falciparum.

XX Key

FH Location/Qualifiers

FT Domain

FT Domain

FT Domain

FT Peptide

FT Domain

FT Domain

FT Domain

FT Peptide

FT Peptide

FT Domain

W09412643-A.

```

PD 09-JUN-1994.
XX 02-DEC-1993: 93WO-A000617.
PF 03-DEC-1992: 92AU-0006206.
PR 16-DEC-1992: 92AU-0006380.
XX (UNIX ) UNISEARCH LTD.
XX Flores MV, Osullivan WJ, Stewart TS:
XX WPI: 1994-200271/24.
XX N-PSDB: AA062924.
XX Nucleic acid encoding carbamoyl phosphate synthetase II -
PI isolated from Plasmodium falciparum, used to develop prods. for
PI the treatment of malaria.
PS Disclosure: Page 6-16; 31pp: English.
XX The cDNA sequence encoding the carbamoyl-phosphate-transferase II
XX (CPSII) of Plasmodium falciparum was determined. The cDNA encodes
XX a protein that includes 2 insert sequences not found in other CPSII
XX proteins. The first separates the putative structural subdomain and
XX the glutamine subdomain of the glutamine-amidotransferase subunit
XX of CPSII, while the second separates 2 ATP binding subdomains of the
XX CPSII subunit, CPSa and CPSb.
SO Sequence 2391 AA:

```

Query Match 3.9% Score 106 DB 15 Length 2391
 Best Local Similarity 20.9% Pred. No. 1.3
 Matches 109: Conservative 80: Mismatches 230: Indels 102: Gaps 25:

```

OY 7 TWYITPLNQ--LKGGTIVNVYV--VKEFKRPPLSKGTDVCSVTIVDQTNWKL-CLL 61
DB 939 smeyeykfrEiaIk:itlnilgecniqfqlnptq---eycl-----enarlsrsa 990
OY 62 FSGNVEALPIIYKNGDIVFH--RLKIOVYKKEGOGITSSGFASLFEGLG--APILP 116
DB 991 laskatgyplayIsakIdIydlIsIknlkkt-----actepIdyIttkip 1040
OY 117 RISSKYFNFT----EDKKVREALRVMASTHNSPWTLLKLDVQPMQYFDLTQQLGKA 172
DB 1041 rwdlnkIefasntmsmksvgeVmsIgrfteesIqksIrcIdnYIqfSntyc----- 1094
OY 173 EVDGASFLKVMQDTRTPFSPSRV--LIQDLVLEGLDSIHRLQMLITIDLVYDNHVA 230
DB 1095 -Idxge--kklceelknpsk-ridaInqafhlmpmkIheltIdyWfl----- 1141
OY 231 RELKVGSLRIYSLHTKLOSMSSENOTMLSEFHLHGSTYGRGIRVLPESNSVDYDLK 290
DB 1142 -----hkfyInIyqnlkIkIkllegIsfndIkYfkKhgfsdKqIahyIsfnsdnnnn 1196
OY 291 DLFsANLTANQSHDYICQSEPDSPSSGSVSLYEVRQQLSATL--LTQHGYL----- 343
DB 1197 nIsScrVlen---dImkyrEklglp-----hIkvIdtIsaefpaltInYltYgqg 1245
OY 344 ERTPLCALIKQKAPQOYRIAPKLSYKPRRLPQSYKVLHCQPKCHLQEVPHNEGDLIIFQD 403
DB 1246 ehndYIplmktk-----kIcIlnkranKkKhv-kihlynevvdKdIdqIbken 1295
OY 404 GAKTIPDVKLQNTSLYDSKIMTTKNOGRKVVNVKNGNGLPLSLNECLLIEGTLSEI 453
DB 1296 nnnn-----mnsqVevKcklnkesygnnsnclnInIenIc-----hdIsIokn 1345
OY 454 CK--LSNKFNSVIPVRSCHDELIDLSAPFLIOGTIHNYG 502
DB 1347 Ik::Innsnsl-----snneve-tnInveSeraShhlyg 1302

```

RESULT 9

AAW06556
 ID AAW06556 standard: Protein: 655 AA.
 XX
 AC AAW06556;
 XX
 DI 12-FEB-1997 (first entry)
 XX
 DE Serine kinase SRPK1.
 XX
 KW Human: serine kinase; SRPK1: SR domain: SR family: splicing factor;
 KW cell proliferation; malignancy; psoriasis; pemphigus vulgaris;
 KW Behcet's syndrome; acute respiratory distress syndrome; leukemia;
 KW rheumatoid arthritis; AIDS; vasculitis; septic shock;
 KW acute inflammation; lipid histiocytosis; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 11..21 "Potential nuclear localisation signal"
 FT 80..100
 FT Domain /note= "Conserved kinase domain I"
 FT 87
 FT Misc-difference /note= "Conserved residue"
 FT 89
 FT Misc-difference /note= "Conserved residue"
 FT 94
 FT Misc-difference /note= "Conserved residue"
 FT 103..119
 FT Domain /note= "Conserved kinase domain II"
 FT 109
 FT Misc-difference /note= "Conserved residue"
 FT 122..129
 FT Domain /note= "Conserved kinase domain III"
 FT 132..146
 FT Domain /note= "Conserved kinase domain IV"
 FT 148..169
 FT Domain /note= "Conserved kinase domain V"
 FT 172..205
 FT Domain /note= "Conserved kinase domain VIa"
 FT 208..225
 FT Domain /note= "Conserved kinase domain VIb"
 FT 213
 FT Misc-difference /note= "Conserved residue"
 FT 215
 FT Misc-difference /note= "Conserved residue"
 FT 218
 FT Misc-difference /note= "Conserved residue"
 FT 257..277
 FT Peptide /note= "Potential nuclear localisation signal"
 FT 492..507
 FT Domain /note= "Conserved kinase domain VII"
 FT 497
 FT Misc-difference /note= "Conserved residue"
 FT 499
 FT Misc-difference /note= "Conserved residue"
 FT 509..523
 FT Domain /note= "Conserved kinase domain VIII"
 FT 520
 FT Misc-difference /note= "Conserved residue"
 FT 521
 FT Misc-difference /note= "Conserved residue"
 FT 526..553
 FT Domain /note= "Conserved kinase domain IX"
 FT 533
 FT Misc-difference /note= "Conserved residue"
 FT 538
 FT Misc-difference /note= "Conserved residue"
 FT 554..573
 FT Domain /note= "Conserved kinase domain X"
 FT 521..554
 FT Domain /note= "Conserved kinase domain XI"
 FT 641
 FT Misc-difference

FT /note= "Conserved residue"

XX US5559019-A.

XX 24-SEP-1996.

XX 22-JUN-1994: 94US-0264002.

XX 22-JUN-1994: 94US-0264002.

PA (REGC) UNIV CALIFORNIA.

PI Fu X, Gui J:

XX WPL 1996-442419/44.

DR N-PSDB: AAT43897.

PI New isolated serine kinase nucleic acid encoding SRPK1 - used to develop prods. which can be used in detection and therapy, partic.

PI for cell proliferative disorders

PS Claim 2: Fig 2: 42pp: English.

CC This sequence represents human serine kinase SRPK1. In the SRPK1
 CC DNA sequence, the sequence preceding the ATG initiation codon contains
 CC stop codons in all three reading frames. The sequence surrounding the
 CC initiation codon conforms to the kozak consensus sequence. SRPK1 acts
 CC to phosphorylate serine residues of the SR domain of the SR family of
 CC splicing factors. The phosphorylation of these sites by SRPK1
 CC affects the ability of the splicing factors to mediate pre-mRNA
 CC splicing and regulates the cellular distribution of splicing factors
 CC during the cell cycle. SRPK1 has a mol. wt. of 92 kD as determined
 CC by SDS-PAGE. The modulation of SRPK1 activity can be used to treat
 CC cell proliferative disorders, e.g. malignancies, psoriasis, pemphigus
 CC vulgaris, Behcet's syndrome, acute respiratory distress syndrome,
 CC leukemia, rheumatoid arthritis, AIDS, vasculitis, septic shock and
 CC other types of acute inflammation, and lipid histiocytosis. The
 CC sequence given in fig. 2 is longer than the sequence given in the
 CC sequence listing in the body of the specification.

SO Sequence 655 AA:

Query Match 3.8%; Score 105; DB 17; Length 655;

Best Local Similarity 20.9%; Pred. No. 0.19; 98; Indels 94; Gaps 13;

Matches 61; Conservative 39; Mismatches 94; Gaps 13;

QY 247 KLGSMSENQIMLSLEFHLHGTS-----YGRGIRVLPESNSDVQDKKDLSEAN---- 296

DB 329 Klessstlgddqlmerdelegaaelncngvievinytqnsnmetlrbkhdlnhandcdv 388

QY 297 LIANQHDVYICQSPDPSFSSGVSILYVERCOQLSATILTDQVLEKPLC---ALIK 353

DB 389 qnlqessfl-----slpqdestsgelcdscplseve-----sdmwcqsstly 434

QY 354 OKAPQGYRIKAKLSYKPRFLFOVKLHCPCXCHLDQVPHGDLII FODGAIKIPDKVL 413

DB 435 qsfceqh-----isqlqesiraelp-edegeqemgprid----- 458

QY 414 CNISLYDSKIWTRNQGKRYAVHFKVNNGLPL--SNECLLLIEGTLSEICKLSKEN 471

DB 459 -----nkxkstaqlnflcn---plepknaeklxkkladlnmacvnhhlt 509

QY 472 SVLPVPSGHEDELLDLSAPFLIOGTIHHVGTGYCTP-----PICVGYDLYT 518

DB 510 editlrfq-yrstetl-----iqsgyntpradwstacmefelat 546

RESULT 10

AA58113
 ID AAY58113 standard: Protein: 1112 AA.
 XX
 AC AAY58113;

```

XX 07-MAR-2000 (first entry)
XX
XX Haemorrhagic enteritis virus (HEV) RNA polymerase.
XX
XX Haemorrhagic enteritis virus (HEV) Immune system: Turkey: sub-nodular
XX Inoculation: haemorrhagic; immunosuppression: RNA polymerase
XX subunit vaccine; antiviral: recombinant; vector: gene therapy
XX
XX Haemorrhagic enteritis virus.
XX
XX MO9960131-A2.
XX
XX 25-NOV-1999.
XX
XX 13-MAY-1999: 99WO-110026R.
XX
XX 20-MAY-1998: 98IL-0124557.
XX
XX (ABIC ) ABIC LTD.
XX
XX Ptkovskij J, Maalem M, Rei Koren Z, Kispel S, Shmueli I;
XX Peretz Y, Gutter B, Gallili G, Michael A, Goldberg D;
XX
XX MPI: 2000-062458/05.
XX N-PSDB: AA245358.
XX
XX New hemorrhagic enteritis virus genes useful as vaccines for treating
XX viral infection in domesticated birds e.g. turkey and in humans
XX
XX Claim 31: Page 63-67: 89pp: English.
XX
XX Sequences AAY58101-58115 represent proteins encoded by the
XX haemorrhagic enteritis virus (HEV). HEV is an adenovirus which infects
XX turkeys, causing intestinal haemorrhaging and immunosuppression.
XX The structural proteins include the 97 kD hexon protein (AAY58107),
XX which is an outer capsid monomer; the 50 kD penton base protein
XX (AAY58103); and the fibre protein (AAY58111), which anchors the penton
XX base protein and plays an important role in the first attachment of the
XX virus to the cell receptor. Other HEV proteins are 52K (AAY58101), 11kA
XX (AAY58102), core protein I (CPI, AAY58104), core protein II (CPII,
XX AAY58105), pVI (AAY58106), endoprotease (EP, AAY58108), 100K
XX (AAY58109), pVII (AAY58110), Iva2 (AAY58112), DNA polymerase (POL,
XX AAY58113), precursor terminal protein (pTP, AAY58114), and DNA binding
XX protein (DBP, AAY58115). These proteins, or the genes encoding them, may
XX be used in the preparation of a subunit vaccine against the virus. Such
XX vaccines are likely to be effective, and also inexpensive, making their
XX use economically worthwhile. Additionally, the fibre protein, which
XX mediates viral attachment, may be modified to alter its host cell
XX specificity. A recombinant HEV may be constructed for use as a vector for
XX gene therapy. The nucleotides encoding the proteins may also be used for
XX diagnostic purposes, or may be used as a source of primers and probes.
XX
XX Sequence 1112 AA:
XX
XX

```

Query Match 3 83: Score 104: DB 21: Length 1112:
 Best Local Similarity 19.08: Pred. No. 0.58:
 Matches 96: Conservative 78: Mismatches 171: Indels 160: Gaps 25:

```

QY 7 TNYIYPLNQLK-----GGTIVNYYGVKFFPKYLSKGDIDYCVIIVDJNWK 56
   : 1111 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11:
DB 2 skyitltgelnelpyrlgickkdkatijefawynlfncckkial-----kvkkaqsyj 53
   : 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11:
QY 57 LTCLFSGNEVALPIIYKNGDIVRFHRL--KIOYKKEITGCI-----TSSGASLTF 105
   : 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11:
DB 54 llsckykpt-----nsdienvknlpvksltlwkklgtgcmqmetkshyfield 104
   : 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11:
QY 107 EGTLCAPLIPPTSSKYFFNTTIEDHKNVEALPYWASIHMSPSITLKLK-DYDPMKYFD- 153
   : 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11:
LH 105 -----lydsksklcklks*vtgqgkoseqrlysaiah-----tcndirstyynsv 149
   : 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11:

```

```

QY 164 -----ETGQILGKAENVKASFLKVMQ-GTRTFPSMRPLIDIVL-----ESDLS 208
   : 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11:
DB 149 dndtktkscitpfpqivmncckklll--lydfoctftlsoosvltmptlltstfnd 204
   : 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11:
QY 209 HIBHGLGHLIDLLVYDNIHIVAKSLKVGSPFLKIVSLHRTKDSMNSNQT-----ML 259
   : 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11:
DB 205 -----dcltctcktdla-----kksktdq kstlywskvkhll 248
   : 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11:
QY 260 SEIIEH-----LHRTSGDGLKRVLPSSNDVQL--KKLES-ANITANSHVICOSE 310
   : 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11:
DB 279 scrttktrnmllfatldyftklillporellfodagqulstlnlnshnmllmslk 298
   : 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11:
QY 311 PHNFPSSGSVSLYVERCCG-----LSATLLTDHLYERTPLCALIKOKAPDOGYRIAR 365
   : 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11:
DB 289 vknltf-----lolytlmnlqsfndllatqldendldplllvernmfmggrtl--- 350
   : 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11:
QY 366 LSVYKRPFLFCGVKMKCP-----KCLIAQF-----PHRSDLDLI FRODQATK 407
   : 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11:
DB 351 -----lndlttkfmpdyvayakeekknkksel lsklealpbhesklnlylksmvr 402
   : 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11:
QY 408 TTVVKLIQHLSTYD-SKIVTTKNQKG 431
   : 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11:
DB 403 -dlqtlstskvokaynkyltkg 426
   : 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11:

RESULT 11
AAB31889
ID AAB31889 standard: Protein: 4393 AA.
AC AAB31889:
XX
XX 15-MAY-2001 (first entry)
XX
XX Amino acid sequence of a human protein.
XX
XX DE
XX
XX KW Human: perlecan: retinol-binding plasma protein: calgranulin B: vaccine:
XX ganglioside GM2 activator: saposin B: degenerative disease: glial cell:
XX neurological disease: auto-immune disease: multiple sclerosis: toxicity:
XX Alzheimer's disease: Parkinson's disease: amyotrophic lateral sclerosis:
XX rheumatoid polyarthritis: lupus erythematosus: gene therapy.
XX
XX OS Homo sapiens.
XX
XX MN W0200105442-A2.
XX
XX PD 25-JAN-2001.
XX
XX XX 17-JUL-2000: 2000WO-FR02057.
XX
XX PF 15-JUL-1999: 99FR-0009372.
XX
XX PE (IMMR ) BIOMERIEUX STELHYS.
XX
XX PA Foecklin D, Kolbe H, Charles M, Marcus C, Santoro L, Perron H;
XX
XX PI MPI: 2001-159475/15.
XX
XX DR Detecting, preventing and treating degenerative, neurological and
XX autoimmune diseases, particularly multiple sclerosis, using specified
XX polypeptides or related nucleic acid or ligand
XX
XX PS Claim 1: Page 138-152: 209pp: French.
XX
XX The present sequence represents a human protein, which is used in the
XX method of the invention. The specification describes a method which uses
XX at least one polypeptide or polynucleotide sequence belonging to the
XX perlecan, precursor of the retinol-binding plasma protein, precursor of
XX the ganglioside GM2 activator, calgranulin B or saposin B protein
XX families. The method is used for detecting, preventing or treating a
XX degenerative, neurological and/or auto-immune disease. The
XX polynucleotides and polypeptides are used for diagnosis, prognosis,
XX prevention and treatment of multiple sclerosis (in its various forms
XX and phases). They may also be useful in cases of e.g. Alzheimer's and

```


CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and
 CC in gene therapy (expression of sense or antisense sequences). They can
 CC also be used to assess efficacy of potential therapeutic agents. They can
 CC particularly compounds that reduce or inhibit toxicity towards glial
 CC cells.

XX Sequence 4393 AA:

Query Match 3.7% Score 102; DB 22; Length 4393;
 Best Local Similarity 24.1% Pred. No. 9.6;
 Matches 81; Conservative 38; Mismatches 137; Indels 8; Gaps 13;

OY 48 TIVDOINWKLICLLFGSGVEALPIIKKNDI-----VFHRLKI-----QVYKKE 92
 Db 2257 tvaegtlidscv-agahagvtykrgslparhvgtrlylfqspadagqvra 2315
 OY 93 TCGITSSGFALTEGTLGA-----PI-IPRTSKTFNFTED----- 125
 Db 2316 snqmeas--ltvtytgltqanlaypagstqptlepsssgyaegqtlidncvvpqgshq 2373
 OY 130 ---HKMEALRWKASTHMSPTLLKLCDOVPMQYFDITCOLGAEDGASFLKXWDG 186
 Db 2374 vtwkrgsgslvrlqthgs-----llrlqaspadagvtycrgslpdeasvltltpa 2429
 OY 147 TRTP----FPMWVLLDPLVLEGDSLHRLQNLITDILVYN-HVHV-----AR 231
 Db 2430 gsvpdlgvtflvrl-----essssgyaegqtlidncivagahagvtykrgslpar 2482
 OY 232 SLKVGSEFRI-----YSLHKLOSMSSENOTMSTLEFHLHGCTSYGPGIRVLEES 281
 Db 2483 hqthgsrlrllyqtpadagvtycrgsvtycgaavltlqrlsgshsgvayprtyies 2542
 OY 282 NSDVOLKKDELSANLTANQSHDVYCOSEPDDSEFS 317
 Db 2543 ssaslaqthlidlncivagahagvtykrgslps 2578

RESULT 12

AAV34460
 ID AAV34460 standard; Protein: 684 AA.

XX AAV34460:

XX 25-AUG-1993 (first entry)

XX Porphyromonas gingivalis protein PG115.

XX Porphyromonas gingivalis; PG: periodontal disease; gingivitis;

XX vaccine; antigenic.

XX Porphyromonas gingivalis.

XX WC9929870-A1.

XX 17-JUN-1999.

XX 10-DEC-1998: 98MO-AUD10123.

XX 04-AUG-1998: 98AU-0005028.

XX 10-DEC-1997: 97AU-0000839.

XX 31-DEC-1997: 97AU-0001182.

XX 30-JAN-1998: 98AU-0001546.

XX 10-MAR-1998: 98AU-0002264.

XX 09-APR-1998: 98AU-0002911.

XX 23-APR-1998: 98AU-0003128.

XX 05-MAY-1998: 98AU-0003338.

XX 22-MAY-1998: 98AU-0003654.

XX 29-JUL-1998: 98AU-0004917.

XX (CSLC-) CSL LTD.

FI Aajus CI, Barr IG, Hocking DM, Margetts MB, Patterson MA;

FI Pos BC, Polhel LJ, Webb EA;

XX WPI: 1999-385613/32.

XX N-PSDE: AAX91578.

XX Antigenic Porphyromonas gingivalis peptides for preventing

XX gingivitis

XX Claim 1: Page 438-439; 588pp; English.

CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
 CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAV34318 to
 CC AAV34583. AAX91802 to AAX91989 represent PCR primers used in the
 CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The PG polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis.

XX Sequence 684 AA:

Query Match 3.5% Score 99; DB 20; Length 684;
 Best Local Similarity 21.1% Pred. No. 0.83;
 Matches 101; Conservative 69; Mismatches 177; Indels 132; Gaps 26;

OY 57 LTCLFGSGVEALPIIKKNDIYVFHRLKIQVY-----KKETOGITSSGFSLTEGTL 110
 Db 270 ltqilly-----fplklnmndi---qtnkqilycngyvtdvqglvpt-fll-lhavt 318
 OY 111 GAPLIPTSSKRYFNFTEDHKMEALRWKASTHMS-----PSWTL 151
 Db 319 dspdlpnrsvryqsdanvki-----sshtlkkvadrlleifkndrptfeekwsl 371
 OY 152 KLCDOVPMQYFDITCOLGAEDGASFLKXWDGTRTPPSRVLIQDVLE----- 204
 Db 372 klf-----veygmldexf--yeraakflltdmgqnhkyltdeyrltvegvqdkdqvyv 425
 OY 205 -----GDSLHRLQNLITDILVYDNHV--HVA-----SLKVGSEFLIYVS----- 243
 Db 426 lyaekkhqgyswhkrasdkysvmldgqlphtlveqlekthfvrvdtdltnlir 485
 OY 244 --LHFKLOSMSSENOTMSTLEFHLHGCTSYGSGIRVLPESNSVDOLKKDLSA--NLTA 299
 Db 486 keeraeckksdteralvkl-fear-----lprde-----khhfnvateslga 527
 OY 300 NQSHDVYCOSE-----PDDSPSSGSVSLYEVEPCOOLSATILTDHOYLERTPLCALIKOK 355
 Db 528 egeaillitgaeftmrmtmdaglqpmstfyg-elpdsynlvintdhphtirv-lsgekesv 585
 OY 356 APOGYRIPAKLRSTYKR--RLFOSVKLHCPKCHLQDEVPEHGDLDIFODGATKTPDVKL 413
 Db 586 epslteltraklaekkaeaklldeekdkp-----eeip-----vatkeak 625
 OY 414 QNTSLYDSKIKWITKQKGRVAVHFNKNGILPLSNCLL--LIEGGLSELCISKNK 459
 Db 627 ennaevagakegeindqltkya---qdnelllqgldalllqsglltgeaiaeftrsq 682

RESULT 13

AAV34336
 ID AAV34336 standard; Protein: 686 AA.

XX AAV34336:

XX 25-AUG-1999 (first entry)

XX Porphyromonas gingivalis protein PG116.

XX Porphyromonas gingivalis; PG: periodontal disease; gingivitis;

XX vaccine; antigenic.


```

Db      2 KIIIIIIIIIIbhsftvsk-----pitelhallslksesftidshspiltskn-1stc 55
Oy      191 FPRSRVLIIDL-----VLEGLSHIHRLQNL-----ID1 220
Db      56 fcsclgtcdvsrlhrvtsldsqhnsqtlssdvahppllpnslaanqisqpiqpjsh 115
Oy      221 LVYDNHVAHVASLKVGEF-----LRIYSLH-----TKLOSMSENOTMISLEFHLH 266
Db      116 lyeirhnljnnnrtngsfpdelssgvlvnlvdlynnnlcqdipvsfntltqlr---hlh 172
Oy      267 GGTSYGFG-----LRLPEESNDVDOLKDLSEANLTLANOHSADVICOSEPPDSF 315
Db      173 lqnyfsgkipatygtvpyvleylavsgneltg-kippeignlttrellygyvnafeqg 231
Oy      316 PSS-GSVSLXVEVERCOCLSAIILIDHOYLERPLCALIKOKAPOQVIRAKLSYKPPRL 374
Db      232 ppeignsls-----239
Oy      375 FOSYKLCRCPKCHLLOEVPHF-----GDLIDIFGD-----GAKTIPDYKLCW 415
Db      240 -elvrifaancqjltgeipeigkqlkldlffq:natfgltlqelqlisslkemdl---s 295
Oy      415 TSLYDSKIMTKKQKGRKVAHFEVKN--GILPLSECLLLEGGTLESEICKL---SKKF 470
Db      296 nmmtgqtpstfsgqlknltllnlfknklygaip---efl-----gempelerqlkxentf 347
Oy      471 NSVIFVPSGHED-LELLDLSAPFLIOGTIMHVGTCVTPIC 511
Db      348 tgsipqklgenqtlvildissnkl-----tqltpmnc 380

RESULT 15
ID      AAB47250
XX      AAB47250 standard: Protein: 653 AA.
AC      AAB47250:
XX      18-JUL-2001 (first entry)
DI      XX
LE      Human P7.
KW      Mitochondrial calcium-binding protein; p95.6/YN52; calcium:
KM      drug screening; sequence extraction; P7; calcium binding motif.
XX      Homo sapiens.
XX      OS
XX      FN      W0200130830-A2.
XX      PD
XX      03-MAY-2001.
XX      FE      26-OCT-2000: 200MWO-US23787.
XX      FP      26-OCT-1999: 59US-0427867.
XX      FR      26-OCT-1999: 59US-0427895.
XX      FR      07-SEP-2000: 200005-0557253.
XX      FA      (MITO-) MITOKOR.
XX      PI      Murphy AN, Fahy ED, Wiley SE, Clevenger W.
XX      DR      WPI: 2001-308615/32.
XX      DR      N-PSDB: AAC85804.
XX      PT      New Caenorhabditis elegans mitochondrial calcium-binding protein,
PI      p95.6/YN52, identified by protein motif database searching, useful for
PT      regulating intracellular calcium ions and in drug screening assays.
XX      PS      Example 2; Fig 13: 176pp: English.
XX      CC      This sequence shows a protein containing calmodulin/calcium
XX      CC      binding motif which was identified using the method of the invention.
XX      CC      The method comprises a computer system for extracting information
XX      CC      from biological sequence databases, which receives a motif that defines

```

```

CC      a desired polypeptide sequence. The computer searches the database
CC      based on a query formulated from the motif to obtain a set of
CC      polypeptide sequences which are then filtered and a report of the
CC      filtered set of polypeptide sequence information is then displayed.
CC      The computer system is useful for identifying recombinant C.elegans
CC      p95.6/YN52 and other mitochondrial proteins involved in binding,
CC      transport or other regulation of intracellular calcium which provides
CC      targets for screening assays that identify agents capable of altering
CC      mitochondrial activity.
XX      -cont-
SQ      Sequence 653 AA:

```

```

Query Match      3.6% Score 98.5; DB 22; Length 653;
Best Local Similarity 19.3%; Pred.No.0.86;
Matches 82; Conservative 51; Mismatches 118; Indels 173; Gaps 21;

```

```

Oy      120 SKYFNFTED-----HKVVALR-----VMASTHNSPMTLLKLCDO-----PMOYFDLTC 166
Db      63 stfslmtenythihkeelelengslseseqdmrdwdydsidydpdsynpqlqplct 119
Oy      167 -----QLGKAEVDGASFLLKWMQGTIRPF-----PSMRVLI-----QD 200
Db      120 didllleatkeqgllhahyvelvfeakvklqmpnfthqtspskevttcqdlnhqlld 179
Oy      201 L-----VLEGL-----SHIHRLQNLITDILVYDNHVAHVAR-----231
Db      180 lflfyknglpsernpyvfnqdfvdrgrkslellmlcstflvypndhlhrgnhedfmm 239
Oy      232 SLKVG---SFLRIYSLHTR--LOSMS-----ENOTMISLEFHLHGTSYGRGI 275
Db      240 nlrygtlkeilnhkylhqrllqllleefyawlpjqtivdnaliv-----lhqgiseltdl 294
Oy      276 RVL-----PESNDVDOLKDLSEANLTLANOHSADVICOSEPPDSFSSGS 320
Db      295 nllhrvcrkmksvllppteirndtdskh-nkygvrfnaghqtkngspelen-----347
Oy      321 VSLYEVEERCOCLSAIILIDHOYLE-----RTPLCAL--LKOKAP 357
Db      348 -----lthexeqjidlwsdprgkngcfnptcrggacyfqpdtaskil 391
Oy      358 GOYRIAPAKLPSY--KPRRLFOYVKLHCPKCHLLOEVPHFEGDLIDIF-----ONGATKTP 409
Db      392 nkygqlkmlirshckpgeq-----yeichdgkvvlflfsasnyeeegsnrga 436
Oy      410 DVKL 413
Db      437 YIKI 440

```

```

Search completed: April 9, 2002, 17:08:16
Job time: 595 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 17:10:41 ; Search time 64.84 Seconds

(without alignments)
767.005 Million cell updates/sec

Title: US-09-816-248-15

Perfect score: 1776
Sequence: 1 MSLVPAINVITYPLNOLKKG.....LRPGSSVARSRLIASTS 340

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_proteol:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1660	93.5	634	4	09NW19 homo sapien
2	1660	93.5	634	4	09NW19 homo sapien
3	348	19.6	384	4	09H662 homo sapien
4	324	18.2	113	11	09CYST7 mus musculu
5	167.5	9.4	555	3	013988 schizosacch
6	153	8.6	463	10	09FNM7 09fmb7 arabidopsis
7	144	8.1	497	5	076380 076380 oxytricha t
8	130	7.3	251	5	045595 045595 caenorhabd1
9	111	6.2	759	3	09C2K8 09c2k8 neotropora
10	109	6.1	587	3	09J3702 09j3702 caenorhabd1
11	103	5.8	1044	4	09Y4E4 09y4e4 homo sapien
12	102.5	5.8	4370	4	09H3V5 09h3v5 homo sapien
13	100.5	5.7	215	5	09XWB2 09xwb2 caenorhabd1
14	98	5.5	1198	10	004538 004538 arabidopsis
15	96	5.4	632	5	015742 015742 dictyostell1
16	96	5.4	752	5	09GUX3 09gux3 dictyostell1
17	95.5	5.4	474	10	049474 049474 arabidopsis
18	94.5	5.3	442	5	09VJ01 09vj01 drosophila
19	94.5	5.3	548	5	09NKD4 09nk4 drosophila

20	92.5	5.2	1124	1	09HK03 09hk03 thermoplasm
21	92	5.2	607	10	09LS89 09ls89 arabidopsis
22	90.5	5.1	610	1	09HPP8 09hpp8 halobacteri
23	90.5	5.1	1030	10	09C905 09c905 arabidopsis
24	90	5.1	361	2	09C125 09c125 lactococcus
25	90	5.1	937	10	09LNX9 09lnx9 arabidopsis
26	89.5	5.0	528	5	016628 016628 caenorhabd1
27	89.5	5.0	667	10	09SGV4 09sgv4 arabidopsis
28	89	5.0	683	2	083436 083436 treponema p
29	89	5.0	1893	1	09P9A8 09p9a8 uncultured
30	89	5.0	3796	6	09ITK4 09itk4 bos taurus
31	88.5	5.0	1217	5	09XV62 09xv62 caenorhabd1
32	88	5.0	3788	11	097412 097412 mus musculu
33	88	5.0	3796	6	097699 097699 bos taurus
34	87.5	4.9	497	13	09LBP8 09lbp8 potamoxyr
35	87.5	4.9	553	3	09HFK5 09hfk5 neotropora
36	87.5	4.9	564	2	053127 053127 mycobacteri
37	87.5	4.9	1406	13	09W6V5 09w6v5 galus gall
38	87	4.9	361	10	09W409 09w409 arabidopsis
39	87	4.9	387	2	09CM28 09cm28 pasteurella
40	87	4.9	484	8	09G4L4 09g4l4 phytomyza i
41	87	4.9	723	10	09SNK7 09snk7 oryza sativ
42	85.5	4.9	318	10	09S137 09s137 arabidopsis
43	85.5	4.9	1002	10	09M221 09m221 arabidopsis
44	85	4.8	175	3	09C2X7 09c2x7 candida alb
45	86	4.8	504	8	09IB28 09ib28 greya punct

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	634 AA.
09NW19	09NW19			
ID	09NW19			
DI	01-OCT-2000 (Trembl)	15, Created		
DI	01-OCT-2000 (Trembl)	15, Last sequence update		
DI	01-OCT-2000 (Trembl)	15, Last annotation update		
DE	CDNA FLJ10368 FIS, CLONE NT2RM2001544.			
OS	Homo sapiens (Human).			
OC	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:			
OC	Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.			
OX	NCRL_TaxID=9605;			
RN	[1]			
RF	SEQUENCE FROM N.A.			
RA	Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,			
RA	Mishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,			
RA	Wakatsuna M., Hosoi T., Kaku Y., Kodaira K., Kondo H., Sugawara M.,			
RA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,			
RA	Matanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,			
RA	Tanamoto J., Nakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,			
RA	Ninomiya K., Iwayanagi T.,			
RI	"NEO human cDNA sequencing project."			
RL	Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AK001230; BAA91568.1; -			
SO	SEQUENCE	634 AA: 71425 MW: C0A9C40823F9540C CRC64:		

Query Match	93.5%	Score 1660:	DB 4:	Length 634:
Best local Similarity	98.8%	Pred. 1e-138:		
Matches	317:	Conservative	2:	Mismatches
			2:	Indels
			0:	Gaps
			0:	
09	1	MSLVPAINVITYPLNOLKKGITVNVGVVFFKPPILSKGTDVCSVYIVDQINWLTCL	60	
DB	1	MSLVPAINVITYPLNOLKKGITVNVGVVFFKPPILSKGTDVCSVYIVDQINWLTCL	60	
09	51	LFSGNVFALPIIYKNGIVFHRKLIOVYKKEIOGIISSGFASLTFEGTIGAPIIRISS	120	
DB	61	LFSGNVFALPIIYKNGIVFHRKLIOVYKKEIOGIISSGFASLTFEGTIGAPIIRISS	120	
09	121	KYFNTFEDHKRWKVALRWMASTHMSPSWTLKLCDOVPMQYFDLTCLGKAEVDCASFL	180	
DB	121	KYFNTFEDHKRWKVALRWMASTHMSPSWTLKLCDOVPMQYFDLTCLGKAEVDCASFL	180	

```

OY 181 LKWDGTRIPFSRWYIIDLVLLEGDLSHIRLONLITIDILVYDNRHVARSLKVGSEFLR 240
    |||||||
DB 181 LKWDGTRIPFSRWYIIDLVLLEGDLSHIRLONLITIDILVYDNRHVARSLKVGSEFLR 240
OY 241 IYSLHTKLOSMNSNQIMLSLEFHLHGCTSYGRCIRVLPESNSDVOLKKDLESANLTAN 300
    |||||||
DB 241 IYSLHTKLOSMNSNQIMLSLEFHLHGCTSYGRCIRVLPESNSDVOLKKDLESANLTAN 300
OY 301 OHSDVICOSEPDSPFPGVSL 321
    |||||||
DB 301 OHSDVICOSEPDSPFPGVSL 321
    |||||||

RESULT 2
O9NMX5 PRELIMINARY: PRT: 634 AA.
AC O9NMX5:
DI 01-OCT-2000 (TREMBlrel. 15, Created)
DI 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE HYPOTHEICAL 71.4 KDA PROTEIN (UNKNOWN) (PROTEIN FOR MGC:1028.).
OS Homo sapiens (Human).
OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
OC Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
OX NCBI_TaxID=9606:
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FLACENTIA:
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki I., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuno Y., Sasaki N.:
RI "NEDO human cDNA sequencing project."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki I., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuno Y., Sasaki N.:
RI "NEDO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=SKIN AND MELANOMA:
RA Strausberg R.:
RA Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK001395: BA931988.1:
DR EMBL: AK022580: BA914110.1:
DR EMBL: BC002923: AA02923.1:
SO SEQUENCE 634 AA: 71441 MW: 123A12CABE708C31 CPG64:

Query Match 93.5%: Score 1660: DB 4: Length 634:
Best Local Similarity 98.8%: Pred. No. 1e-139:
Matches 317: Conservative 2: Mismatches 2: Indels 0: Gaps 0:

OY 1 MSIVPAIVYITPTNOLKGIIVWYGVKFFRPYLSKGIYDYSVYIVQIINWKL 60
    |||||||
DB 1 MSIVPAIVYITPTNOLKGIIVWYGVKFFRPYLSKGIYDYSVYIVQIINWKL 60
OY 61 LFSGNVEALPIYKNGDIVRHRPKIOVKKETOGITSSGFASLTFFESLTGAPIIPIISS 120
    |||||||
DB 61 LFSGNVEALPIYKNGDIVRHRPKIOVKKETOGITSSGFASLTFFESLTGAPIIPIISS 120
OY 121 KYFNFTIEDHVEALPYWASTHMSPSWTLKLCDOVPMQYFDITCOLLGAEDVGSFTL 190
    |||||||
DB 121 KYFNFTIEDHVEALPYWASTHMSPSWTLKLCDOVPMQYFDITCOLLGAEDVGSFTL 190
OY 181 LKWDGTRIPFSRWYIIDLVLLEGDLSHIRLONLITIDILVYDNRHVARSLKVGSEFLR 240
    |||||||
DB 181 LKWDGTRIPFSRWYIIDLVLLEGDLSHIRLONLITIDILVYDNRHVARSLKVGSEFLR 240

```

```

OY 241 IYSLHTKLOSMNSNQIMLSLEFHLHGCTSYGRCIRVLPESNSDVOLKKDLESANLTAN 300
    |||||||
DB 241 IYSLHTKLOSMNSNQIMLSLEFHLHGCTSYGRCIRVLPESNSDVOLKKDLESANLTAN 300
OY 301 OHSDVICOSEPDSPFPGVSL 321
    |||||||
DB 301 OHSDVICOSEPDSPFPGVSL 321
    |||||||

RESULT 3
O9H652 PRELIMINARY: PRT: 384 AA.
AC O9H652:
DI 01-MAR-2001 (TREMBlrel. 16, Created)
DI 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE CDNA: FLJ22581 FIS, CLONE HS102573.
OS Homo sapiens (Human).
OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
OC Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
OX YEBL_TaxID=9606:
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HUMAN SMALL INTESTINE:
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Oobayashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isegai T., Sugano S.:
RI "NEDO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK025234: BA915404.1:
SO SEQUENCE 384 AA: 43151 MW: 6CC968D595CECAC6 CRC64:

Query Match 19.6%: Score 348: DB 4: Length 384:
Best Local Similarity 94.4%: Pred. No. 1.1e-22:
Matches 67: Conservative 2: Mismatches 2: Indels 0: Gaps 0:

OY 251 MNSNDIMLSLEFHLHGCTSYGRCIRVLPESNSDVOLKKDLESANLTANOHSDVICOSE 310
    |||||||
DB 1 MNSNDIMLSLEFHLHGCTSYGRCIRVLPESNSDVOLKKDLESANLTANOHSDVICOSE 310
OY 311 PDSPFPGVSL 321
    |||||||
DB 61 PDSPFPGVSL 71

RESULT 4
O9CYS7 PRELIMINARY: PRT: 113 AA.
AC O9CYS7:
DI 01-JUN-2001 (TREMBlrel. 17, Created)
DI 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 2810458H16RIK PROTEIN.
GN 2810458H16RIK.
OS Mus musculus (Mouse).
OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
OC Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
OX NCBI_TaxID=10090:
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO:
RA STFAIN-C57BL/6J: TISSUE=EMBRYO:
RA MECLINE-21085560: PubMed:11217851:
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Iizaka M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Flechterman W., Gaasterland T., Gissi C., King B., Kochiya H.,
RA Koehl P., Lewis S., Matsuo S., Nikaide I., Pesole G., Quackenbush J.,
RA Schrijm L.M., Stabli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 17:10:44 ; Search time 54.84 Seconds
(without alignments)

1168.554 Million cell updates/sec

Title: US-09-816-248-17

Sequence: 1 MSLVPAINVITPLNOLKGG.....HHYGTGCTPPIVCYDLYT 518

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2637	96.5	634	4	Q9NM19
2	2628	96.2	634	4	Q9NM19
3	1371	48.4	384	4	Q9NM62
4	976	35.7	317	4	Q9NM18
5	324	11.9	113	11	Q9C1S7
6	201.5	7.4	555	3	Q13988
7	153.5	5.6	463	10	Q9FNH7
8	144.5	5.3	497	5	Q76380
9	130	4.8	587	5	Q93702
10	117.5	4.3	251	5	Q45595
11	115	4.2	115	12	Q9QMH2
12	111	4.1	759	3	Q9C1K8
13	110.5	4.0	483	10	Q9XE80
14	110.5	4.0	1044	4	Q9Y4E4
15	108	4.0	1217	5	Q9XV62
16	106	3.9	2391	5	Q27732
17	105.5	3.9	548	5	Q9AKD4
18	105	3.8	655	4	Q12890
19	104	3.8	1112	12	Q9YUR6

20	103.5	3.8	493	2	Q9JMH8	Q9JMH8 wolbachia s
21	103.5	3.8	892	10	Q9XEJ5	Q9XEJ5 zea mays (m
22	103	3.8	896	2	Q9GJ12	Q9GJ12 lactococcus
23	103	3.8	1434	5	Q62239	Q62239 caenorhabdi
24	102.5	3.8	520	13	Q91R52	Q91R52 xenopus lae
25	102.5	3.8	528	5	Q16528	Q16528 caenorhabdi
26	102.5	3.8	698	10	Q9SLM4	Q9SLM4 arabidopsis
27	102	3.7	4370	4	Q9H3V5	Q9H3V5 homo sapien
28	101.5	3.7	737	10	Q9SJ13	Q9SJ13 arabidopsis
29	101.5	3.7	1038	2	Q99R12	Q99R12 staphylococ
30	101	3.7	442	5	Q9VJU1	Q9VJU1 drosophila
31	101	3.7	773	3	Q9SHP2	Q9SHP2 arabidopsis
32	100.5	3.7	1438	3	Q06681	Q06681 saccharomyc
33	100	3.7	1110	10	Q9F177	Q9F177 arabidopsis
34	99	3.6	684	2	Q9S3C2	Q9S3C2 porphyromon
35	99	3.6	712	8	Q19880	Q19880 cyanidium c
36	99	3.6	1002	10	Q9M221	Q9M221 arabidopsis
37	99	3.6	1428	5	Q77023	Q77023 drosophila
38	98.5	3.6	735	5	Q9N8A6	Q9N8A6 trypanosoma
39	98.5	3.6	1003	5	Q9TXJ6	Q9TXJ6 leishmania
40	98.5	3.6	1619	5	Q77382	Q77382 plasmodium
41	98	3.6	909	5	Q17429	Q17429 caenorhabdi
42	98	3.6	2747	5	Q9BJX9	Q9BJX9 plasmodium
43	97.5	3.6	755	12	Q56271	Q56271 human herpes
44	97	3.6	372	2	Q24967	Q24967 helicobacte
45	97	3.6	1123	10	Q9SE19	Q9SE19 arabidopsis

ALIGNMENTS

RESULT 1
Q9NM19
AC Q9NM19
DI 01-OCT-2000 (TrEMBLrel. 15, last sequence update)
DI 01-JUN-2001 (TrEMBLrel. 17, last annotation update)
DE HYPOTHETICAL 71.4 KDA PROTEIN (UNKNOWN) (PROTEIN FOR MGC:10280).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Iisogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuno Y., Sasaki N.,
RA *RED0 human cDNA sequencing project.*
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Iisogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuno Y., Sasaki N.,
RA *RED0 human cDNA sequencing project.*
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP TISSUE=SKIN, AND MELANOMA.
RA Strausberg R.,
RA Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK001935; BAA91988.1;
DR EMBL: AK022580; BAB14110.1;
DR EMBL: BC002923; AAH02923.1;
SQ SEQUENCE 634 AA: 71441 MW: 123A12CABE708C91 CRC64:

Query Match 96.5% Score 2637; DB 4; Length 634;
Best Local Similarity 99.2% Pred. No. 1.2e-214;


```

OY 181 LKXWDGRTFFECNRYLIODLVEGDLNHRLOHLITDILVYDNHNVABSLKVGSLR 240
Db 181 LKXWDGRTFFECNRYLIODLVEGDLNHRLOHLITDILVYDNHNVABSLKVGSLR 240
OY 241 IYSLHTILOSMSNEOTMTLSEFHLHGSTSYGRGIRVLPESNSVDYCIKKOLESANLTAN 300
Db 241 IYSLHTILOSMSNEOTMTLSEFHLHGSTSYGRGIRVLPESNSVDYCIKKOLESANLTAN 300
OY 301 OHSVYICGSEPDSEPFSSGVSLSYEYERCOOLATILTDHGYLFEFFCALIKKAPQOY 360
Db 301 OHSVYICGSEPDSEPFSSGVSLSYEYERCOOLATILTDHGYLFEFFCALIKKAPQOY 360
OY 361 RIRAKLSYKPRRLFSOVKHLCPKRCCHLOEVPHEGDLIIIFODGAIIPVYKLOMTSLYD 420
Db 361 RIRAKLSYKPRRLFSOVKHLCPKRCCHLOEVPHEGDLIIIFODGAIIPVYKLOMTSLYD 420
OY 421 SKIITRNKGKPVAAVHFFVKNGLIPLSNHCILLIEGGLISFCYLSKPSVNPVREGRH 480
Db 421 SKIITRNKGKPVAAVHFFVKNGLIPLSNHCILLIEGGLISFCYLSKPSVNPVREGRH 480
OY 481 EDLELIDLASFILOGCIHHYGGKCCSSLPSTIONLSLYDKTSWIPSSVAEALGIVPELOY 540
Db 481 EDLELIDLASFILOGCIHHYGGKCCSSLPSTIONLSLYDKTSWIPSSVAEALGIVPELOY 540
OY 541 VFVWTFILDDGTYGLEVLYKEDSKFFOIPASEVLMDDIIOKVMDHNMFCPPGJIKIDAY 600
Db 541 VFVWTFILDDGTYGLEVLYKEDSKFFOIPASEVLMDDIIOKVMDHNMFCPPGJIKIDAY 600
OY 601 FWLECFIKSYNVTNGIDNOICYOIFDTTVAEDVI 634
Db 601 FWLECFIKSYNVTNGIDNOICYOIFDTTVAEDVI 634

RESULT 2
O9NMXX5 PRELIMINARY: PRT: 634 AA.
ID 09NMXX5:
AC 09NMXX5:
AD 01-OCT-2000 (IREMBLrel. 15, Created)
DI 01-OCT-2000 (IREMBLrel. 15, Last sequence update)
DI 01-JUN-2001 (IREMBLrel. 17, Last annotation update)
DE HYPOHEMICAL 71.4 KDA PROTEIN (UNKNOWN) (PROTEIN FOR MSG:10280).
OC Homo sapiens (Human)
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX MCLTAXID=9606;
RN 11)
RP SEQUENCE FROM N.A.
PC TISSUE=PLACENTA;
PA Iscovai T., Ota T., Hayashi K., Sugiyama T., Otsuki I., Suzuki Y.,
PA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
PA Matsunaga H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
PA Nakamura Y., Nagahara K., Masuno Y., Sasaki H.
PA *NEBO human cDNA sequencing project.*
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
PL 12)
RN 13)
RP SEQUENCE FROM N.A.
PC TISSUE=SKIN, AND MELANOMA;
PA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK001935; BAA01988.1;
DR EMBL: AK022580; BAB14110.1;
DR EMBL: BC002923; AAH02923.1;
DR EMBL: 634 AA; 71441 MW; 123A12CABE70D9C91 CRC6;;
SEQUENCE

```

Query Match	99.8%	Score 3317	DB 4	Length 634
Best Local Similarity	99.8%	Pred. No. 4,2e-265		
Matches 633	Conservative	0	Mismatches 1	Indels 0
				Gaps 0
Oy	1	MSLVPAITNTITPLNOLKGGITVNVYGVYKFRFPYLSKTDYGVYIVYDONTNKLCTL	60	
Db	1	MSLVPAITNTITPLNOLKGGITVNVYGVYKFRFPYLSKTDYGVYIVYDONTNKLCTL	60	
Oy	61	LESGNEALPIIYKNGDIIVREHLKIQYKKETIGITSSGFASLFEETLGAPIIPRTSS	120	
Db	61	LESGNEALPIIYKNGDIIVREHLKIQYKKETIGITSSGFASLFEETLGAPIIPRTSS	120	
Oy	121	KTFNTTEHKVVEALRWASTHMSPSNTLKLCDVOPMOTFDLTOLLGKAEVDASFLR	180	
Db	121	KTFNTTEHKVVEALRWASTHMSPSNTLKLCDVOPMOTFDLTOLLGKAEVDASFLR	180	
Oy	181	LKVMDSITPIPFSSKRWLLIODLVLEEDLSHIRLONLIDILVYDNVHVARSLKVGSEFLR	240	
Db	181	LKVMDSITPIPFSSKRWLLIODLVLEEDLSHIRLONLIDILVYDNVHVARSLKVGSEFLR	240	
Oy	241	IYSLHKLKLSNMSSENOTMLSTLEFHLHGISTVGRGRVLPENSDVDOLKDELSANTLN	300	
Db	241	IYSLHKLKLSNMSSENOTMLSTLEFHLHGISTVGRGRVLPENSDVDOLKDELSANTLN	300	
Oy	301	QHSDDVJCSSEPPDSFSSGSGSLYEVEERCOOLSATIITLDHOLYERTPLCALIKOKAPOY	360	
Db	301	QHSDDVJCSSEPPDSFSSGSGSLYEVEERCOOLSATIITLDHOLYERTPLCALIKOKAPOY	360	
Oy	361	RIRAKLRKRPPLFOSVYKHCPRKHLLOEYHEDDLDIIFODATITPVYKIONTSYLD	420	
Db	361	RIRAKLRKRPPLFOSVYKHCPRKHLLOEYHEDDLDIIFODATITPVYKIONTSYLD	420	
Oy	421	SKITWIKNGKRVAVHFKVNGILPLSNECLLIEGISTELCKLSNKNYSYIVPVSCH	480	
Db	421	SKITWIKNGKRVAVHFKVNGILPLSNECLLIEGISTELCKLSNKNYSYIVPVSCH	480	
Oy	481	EDLELDLSAPFLDGTTHHGYCKOCSSLRSTONLSYDTSWIPSSVEALGIVPLQY	540	
Db	481	EDLELDLSAPFLDGTTHHGYCKOCSSLRSTONLSYDTSWIPSSVEALGIVPLQY	540	
Oy	541	VFWYTFILDVGIVLEAVYLMDSKRFQIPASEVCLMDDDLOKSVIMDMCPGKIDAY	600	
Db	541	VFWYTFILDVGIVLEAVYLMDSKRFQIPASEVCLMDDDLOKSVIMDMCPGKIDAY	600	
Oy	601	PWLECFIKSYNTNGIDNOICYOIFDTIYAEVI	634	
Db	601	PWLECFIKSYNTNGIDNOICYOIFDTIYAEVI	634	
RESULT	3			
ID	09H662	FELLMINARY:	PRI:	384 AA.
AC	09H662			
DT	01-MAR-2001 (TEMBLrel. 15, Created)			
DT	01-MAR-2001 (TEMBLrel. 15, Last sequence update)			
DT	01-MAR-2001 (TEMBLrel. 15, Last annotation update)			
DE	CDBA: FJ22581 F15, CLONE HS10573.			
OS	Homo sapiens (human).			
OC	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:			
CC	Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.			
OX	NCBI_Taxid=9606;			
FN	(1)			
RP	SEQUENCE FROM H.A.			
RC	TISSUE=HUMAN SMALL INTESTINE.			
FA	Watanabe K., Kimagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,			
FA	Suzuki Y., Okajashi M., Nishi T., Shibahara T., Tanaka T.,			
FA	Nakamura Y., Isigai T., Sugano S.,			
FI	REDU human cDNA sequencing project.			
EL	Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.			
EMBL	AK026234: BAB15404.1			
SD	SEQUENCE 384 AA: 42161 MW: 66C968D5959CEACAC CAC64:			

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 17:06:41 ; Search time 64.84 Seconds
(without alignments)
1430.238 Million cell updates/sec

Title: US-09-816-248-13

Sequence: 1 MSLVPAIRNYITPLNOLKGG.....GIDNOICVQIDTIVAEVDI 634

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: SPREMBL_17:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.mhc:*
9: sp.organelle:*
10: sp.phage:*
11: sp.plant:*
12: sp.rodent:*
13: sp.virus:*
14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3324	100.0	634	4	Q9NM19
2	3317	99.8	634	4	Q9NM19
3	2001	50.2	384	4	Q9H652
4	1655	49.8	317	4	Q9S018
5	716	21.5	135	4	Q9UC95
6	324	9.7	113	11	Q9CY57
7	199.5	6.0	453	10	Q9FNH7
8	153.5	4.3	497	5	Q76380
9	144.5	4.3	587	5	Q93702
10	130.5	3.9	587	5	Q93702
11	123	3.7	115	12	Q9GMH2
12	122	3.7	114	10	Q40392
13	118	3.5	1217	5	Q9XV62
14	117.5	3.5	251	5	Q45595
15	116	3.5	1044	4	Q9V4E4
16	114.5	3.4	745	3	Q9P7X5
17	114.5	3.4	1110	10	Q9T177
18	114	3.4	899	10	Q9SG75
19	114	3.4	917	10	Q9CAF6

20	114	3.4	1348	5	Q9VAD1
21	114	3.4	1475	10	Q9SXQ4
22	114	3.4	2391	5	Q27732
23	112.5	3.4	1112	10	Q9S752
24	112	3.4	907	10	Q9FW49
25	112	3.4	1421	10	Q9SX05
26	112	3.4	2528	5	Q19317
27	111.5	3.4	483	10	Q9XE80
28	111	3.3	759	3	Q9C2K8
29	111	3.3	1447	10	Q9SX03
30	110	3.3	896	2	Q9CGJ2
31	109.5	3.3	1456	10	Q9ZT94
32	109	3.3	1534	10	Q9FZK7
33	108.5	3.3	892	10	Q9XEJ5
34	108	3.2	927	2	Q31052
35	107.5	3.2	980	10	Q9ZU10
36	107.5	3.2	1784	10	Q9C6R1
37	107	3.2	501	1	Q26978
38	106	3.2	773	10	Q9SHP2
39	106	3.2	1123	10	Q9SSL9
40	106	3.2	1265	5	Q9VBR8
41	105.5	3.2	548	5	Q9NRD4
42	105.5	3.2	1038	2	Q99RD2
43	105.5	3.2	6420	2	P95814
44	105	3.2	695	5	Q9VSN5
45	105	3.2	1112	12	Q9YUR6

ALIGNMENTS

RESULT	ID	Q9NM19	PRELIMINARY:	PRI:	634 AA.
AC	Q9NM19:				
DT	01-OCT-2000 (TRENBLREL. 15, Created)				
DT	01-OCT-2000 (TRENBLREL. 15, Last sequence update)				
DT	01-OCT-2000 (TRENBLREL. 15, Last annotation update)				
DE	CDNA FL10368 FIS, CLONE NTZRM2001544.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Isoqal T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,				
RA	Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,				
RA	Wagatsuna M., Hosokiri T., Raku Y., Kodaira H., Kondo H., Sugawara M.,				
RA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,				
RA	Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,				
RA	Yamanoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y.,				
RA	Ninomiyu K., Iwayanagi T.,				
RI	"NEBD human cDNA sequencing project."				
RL	Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL: AK001230; BAA91568.1;				
SO	SEQUENCE 634 AA: 71425 MW: C0A9C4082F9540C CRC64:				

Query Match	100.0%	Score 3324:	DB 4:	Length 634:
Best local Similarity	100.0%:	Pred. No. 1:	le-265:	
Matches 634:	Conservative	0:	Mismatches	0:
			Indels	0:
			Gaps	0:
QY	1	MSLVPAIRNYITPLNOLKGGIIVNYGVKFRPKPLSGDYGSVVIIVDQINVKLCL	60	
DB	1	MSLVPAIRNYITPLNOLKGGIIVNYGVKFRPKPLSGDYGSVVIIVDQINVKLCL	60	
QY	51	LFSGNNEALPIIYKNDIVFRHLKIQVYKKETOGITSSGFASLTFFETGLGAPILPRTS	120	
DB	51	LFSGNNEALPIIYKNDIVFRHLKIQVYKKETOGITSSGFASLTFFETGLGAPILPRTS	120	
QY	121	KTFNFTEDHKWEALRWASTHMSPSMTLLKICVOPQYFDLTCLLGAKVDSASFL	180	
DB	121	KTFNFTEDHKWEALRWASTHMSPSMTLLKICVOPQYFDLTCLLGAKVDSASFL	180	

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 12, 2002, 08:33:16 ; Search time 22.44 Seconds

(without alignments)
710.503 Million cell updates/sec

Title: US-09-816-248-5
Sequence: 1 MSYPAINIVYITPLNOLKGG.....KKETOGITSSGFASLTFT 109

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SPREMBL_17:*
2: sp_archea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhch:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_rodent:*
13: sp_virus:*
14: sp_vertebrate:*
15: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	567	100.0	634	4 Q9NM19	Q9NM19 homo sapien
2	567	100.0	634	4 Q9NM19	Q9NM19 homo sapien
3	324	57.1	113	11 Q9C1S7	Q9C1S7 mus musculu
4	114	20.1	555	3 Q1398E	Q1398E schizosacch
5	106	18.7	497	5 Q76380	Q76380 oxytricha t
6	85	15.0	318	10 Q9S137	Q9S137 arabidopsis
7	77.5	13.7	473	5 Q9Y119	Q9Y119 cryptospori
8	75	13.2	759	3 Q9C2K8	Q9C2K8 neurospora
9	74.5	13.1	383	11 Q9C0E7	Q9C0E7 mus musculu
10	74	13.1	198	4 Q9S330	Q9S330 homo sapien
11	74	13.1	228	4 Q9B0N7	Q9B0N7 homo sapien
12	74	13.1	345	4 Q9P1L3	Q9P1L3 homo sapien
13	74	13.1	383	4 Q9Y282	Q9Y282 homo sapien
14	72.5	12.8	299	5 Q9NF41	Q9NF41 drosophila
15	72.5	12.8	937	10 Q9LNX9	Q9LNX9 arabidopsis
16	71.5	12.6	1054	2 Q9KRF1	Q9KRF1 bacillus ha
17	70.5	12.5	247	2 Q9ZFO1	Q9ZFO1 buchnera ap
18	70.5	12.4	287	2 Q46329	Q46329 cowdria rum
19	70.5	12.4	287	2 Q46331	Q46331 cowdria rum

20	70.5	12.4	287	2 Q9R425	Q9R425 cowdria rum
21	69.5	12.3	313	2 Q9AKM8	Q9AKM8 rickettsia
22	69.5	12.3	313	2 Q9AKH3	Q9AKH3 rickettsia
23	69.5	12.3	383	2 P74756	P74756 synchocyst
24	69.5	12.3	457	2 Q34798	Q34798 bacillus su
25	69.5	12.3	886	10 Q9Z007	Q9Z007 arabidopsis
26	68.5	12.1	812	2 Q9RTS3	Q9RTS3 delnoccocus
27	68.5	12.1	1124	1 Q9HK03	Q9HK03 thermoplasma
28	68.5	12.0	792	3 Q9HC45	Q9HC45 klyveromyc
29	67.5	11.9	306	2 Q9KRF9	Q9KRF9 bacillus ha
30	67.5	11.9	877	12 Q99BV2	Q99BV2 baboon cito
31	67	11.8	480	5 Q17840	Q17840 caenorhabdi
32	67	11.8	497	2 Q9K3R9	Q9K3R9 streptomyces
33	67	11.8	518	5 P91345	P91345 caenorhabdi
34	67	11.8	1175	2 Q47955	Q47955 haemophilus
35	66.5	11.7	490	5 Q22200	Q22200 caenorhabdi
36	66.5	11.7	664	3 Q9URD2	Q9URD2 schizosacch
37	66.5	11.7	801	3 Q9P3N3	Q9P3N3 schizosacch
38	66	11.6	279	13 Q9PUS5	Q9PUS5 brachydanio
39	66	11.6	576	10 Q9LIR8	Q9LIR8 plasmodium
40	65	11.6	2867	5 Q9N2M3	Q9N2M3 plasmodium
41	65.5	11.5	284	2 Q9AF41	Q9AF41 cowdria rum
42	65.5	11.6	313	2 Q9AKC5	Q9AKC5 rickettsia
43	65.5	11.6	598	12 Q9YW15	Q9YW15 melanoplus
44	65	11.5	251	2 Q9ZEY5	Q9ZEY5 buchnera ap
45	65	11.5	279	2 Q9RC44	Q9RC44 streptococc

ALIGNMENTS

RESULT 1	
Q9NM19	PRELIMINARY: PRT: 634 AA.
AC Q9NM19;	
DI 01-OCT-2000 (TrEMBLrel. 15, Created)	
DI 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)	
DE CDNA FLJ10368 FIS. CLONE NT2RM2001544.	
OS Homo sapiens (Human).	
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC Mammalia; Eultheria; Primates; Catarrhini; Homnidae; Homo.	
OX NCBI_TaxID:9606;	
PP [1]	
FP SEQUENCE FROM N.A.	
PA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,	
PA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,	
PA Watanabe S., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,	
PA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,	
PA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuo Y.,	
PA Nimomiya K., Iwayanagi T.,	
RI "MED" human cDNA sequencing project."	
RI Submitted (FEF-2000) to the EMBL/GenBank/DBJ databases.	
DR EMBL AK001230; BAA91568.1; -	
SQ SEQUENCE 534 AA: 71425 MW: C0A9C40823F9540C CRC64:	

Query Match	100.0%; Score 567; DB 4; Length 634;
Best local Similarity	100.0%; Pred. No. 1; le-51;
Matches 199:	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MSYPAINIVYITPLNOLKGGTIVNVYGVKFFKPYLSKGTGDSVTVTDOTNNKLTCL 60
DB	1 MSYPAINIVYITPLNOLKGGTIVNVYGVKFFKPYLSKGTGDSVTVTDOTNNKLTCL 60
QY	61 LFGSGEALPIIYKNGDVFHRLKIOVKKETOGITSSGFASLTFT 109
DB	61 LFGSGEALPIIYKNGDVFHRLKIOVKKETOGITSSGFASLTFT 109
RESULT 2	
Q9NM19	

```

ID 09NMUX5 PRELIMINARY: PRT: 634 AA.
AC 09NMUX5:
DI 01-OCT-2000 (IREMBLrel. 15, Created)
DI 01-OCT-2000 (IREMBLrel. 15, Last sequence update)
DI 01-JUN-2001 (IREMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 71.4 KDA PROTEIN (UNKNOWN) (PROTEIN FCP MGC:10290).
OS Homo sapiens (human).
OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
OC Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
OX NCBI_TaxID=9606;
PN 11
PP SEQUENCE FROM N.A.
PC TISSUE=PLACENTA.
RA Isegaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki H.,
RA "MEDO human cDNA sequencing project."
RI Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RL 12
PP SEQUENCE FROM N.A.
RA Isegaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki H.,
RA "MEDO human cDNA sequencing project."
RI Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RL 13
PP SEQUENCE FROM N.A.
PC TISSUE=SKIN, AND MELANOMA.
RA Strausberg R.,
RA Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK001935: BAA91988.1: -
DR EMBL: AK022580: BAB14110.1: -
DR EMBL: BC002923: AA02923.1: -
SQ SEQUENCE 634 AA: 71441 MW: 123A12CABE70BC91 CFC64:

Query Match 100.0% Score 567: DB 4: Length 634:
Best Local Similarity 100.0%: Pred. No. 1, 1e-51:
Matches 109: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

OY 1 MSLVPATNYYTLPNLQKGGIIVNVGVKFFRPYLSKGTGYCSVTIVDQTNVKLTCL 60
DI 1 MSLVPATNYYTLPNLQKGGIIVNVGVKFFRPYLSKGTGYCSVTIVDQTNVKLTCL 60
DB 1 MSLVPATNYYTLPNLQKGGIIVNVGVKFFRPYLSKGTGYCSVTIVDQTNVKLTCL 60
OY 61 LFSGNYEALPIIYKNGDIVFHFRLKIOVYKKEITSSGFASITFEET 109
DI 61 LFSGNYEALPIIYKNGDIVFHFRLKIOVYKKEITSSGFASITFEET 109
DB 61 LFSGNYEALPIIYKNGDIVFHFRLKIOVYKKEITSSGFASITFEET 109

RESULT 3 PRELIMINARY: PRT: 113 AA.
OY 09CYS7:
DI 01-JUN-2001 (IREMBLrel. 17, Created)
DI 01-JUN-2001 (IREMBLrel. 17, Last sequence update)
DI 01-JUN-2001 (IREMBLrel. 17, Last annotation update)
DE 2810458H15R1K PROTEIN.
GN 2810458H15R1K.
OS Mus musculus (mouse).
OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
OC Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
OX NCBI_TaxID=10090;
PN 11
PP SEQUENCE FROM N.A.
PC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085650; PubMed=11217891;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosaka H., Kondo S., Yamana K. I.,
RA Saito K., Okazaki Y., Gotohori T., Bono H., Kasuoka T., Saito R.,
RA Saito K., Matsuda H. A., Ashburner S., Balaban S., Casavant T.,

```

```

FA Fleischiemann W., Gaasterland T., Gissi C., King B., Koichiwa H.,
FA Kuehl P., Lewis S., Matsuo Y., Nikaide J., Pesole G., Quackenbush J.,
FA Schriml L. M., Stambli F., Suzuki R., Tomita M., Wagner L., Washio T.,
FA Sakai K., Okido T., Furuno M., Aono H., Balaban R., Barsh G.,
FA Blake J., Boffelli D., Boujarda N., Carninci P., de Bonaldo M. F.,
FA Broststein M. J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
FA Gustincich S., Hill D., Hofmann M., Hume D. A., Kamiya M., Lee N. H.,
FA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,
FA Nordone P., Ping B., Ringwald M., Rodriguez I., Sakamoto N.,
FA Sasaki H., Sato K., Schenbach C., Segal T., Shibata Y., Storch K.-F.,
FA Suzuki H., Togo-Oka K., Wang K. H., Wetz C., Whitaker C., Wilming L.,
FA Wyshnyavskiy A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,
FA Hasehizaki Y.,
FI Functional annotation of a full-length mouse cDNA collection."
PL Mature 493,685-690(2001).
DR EMBL: AK013364: BAB28810.1: -
DR MGD: MGI:1920085: 2810458H15R1K.
SQ SEQUENCE 113 AA: 12740 MW: 08BB9FC5D0C7088 CFC64:

Query Match 57.1% Score 324: DB 11: Length 113:
Best Local Similarity 73.3%: Pred. No. 7, 2e-27:
Matches 63: Conservative 6: Mismatches 17: Indels 0: Gaps 0:

OY 1 MSLVPATNYYTLPNLQKGGIIVNVGVKFFRPYLSKGTGYCSVTIVDQTNVKLTCL 60
DI 1 MSLVPATNYYTLPNLQKGGIIVNVGVKFFRPYLSKGTGYCSVTIVDQTNVKLTCL 60
DB 1 MSLVPATNYYTLPNLQKGGIIVNVGVKFFRPYLSKGTGYCSVTIVDQTNVKLTCL 60
OY 61 LFSGNYEALPIIYKNGDIVFHFRLKIOVYKKEITSSGFASITFEET 109
DI 61 LFSGNYEALPIIYKNGDIVFHFRLKIOVYKKEITSSGFASITFEET 109
DB 61 LFSGNYEALPIIYKNGDIVFHFRLKIOVYKKEITSSGFASITFEET 109

RESULT 4 PRELIMINARY: PRT: 555 AA.
OY 013988:
DI 01-JAN-1999 (IREMBLrel. 09, Created)
DI 01-MAY-2000 (IREMBLrel. 13, Last sequence update)
DI 01-MAY-2000 (IREMBLrel. 13, Last annotation update)
DE HYPOTHETICAL 64.1 KDA PROTEIN C26H5.06 IN CHROMOSOME 1.
GN SPAC26H5.06.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota: Fungi: Ascomycota: Schizosaccharomycetes:
OC Schizosaccharomycetales: Schizosaccharomycetaceae.
OX Schizosaccharomyces.
OX NCBI_TaxID=4896;
PN 11
PP SEQUENCE FROM N.A.
PC STRAIN=972;
RA Oliver K., Harris D., Barrell B. G., Rajandream M. A., Wood V.,
RA Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RI EMBL: Z5126: CAB16192.2: -
RI Hypothetical protein.
SQ SEQUENCE 555 AA: 64111 MW: A79DNA95A0C4F803 CFC64:

Query Match 20.1% Score 114: DB 3: Length 555:
Best Local Similarity 34.4%: Pred. No. 0, 00074:
Matches 31: Conservative 18: Mismatches 35: Indels 5: Gaps 3:

OY 18 KEGIIVNVGVKFFRPYLS-KGT-DYCSVTIVDQ-----NKLITLFSGNYEALPI 71
DI 1 KEGIIVNVGVKFFRPYLS-KGT-DYCSVTIVDQ-----NKLITLFSGNYEALPI 71
DB 1 KEGIIVNVGVKFFRPYLS-KGT-DYCSVTIVDQ-----NKLITLFSGNYEALPI 71
OY 38 KKHIVNLCGVKIDFTPSQSLHGTAKDWTIVYIMDPICDTSISGLQIHLSKQNDLPV 97
DI 38 KKHIVNLCGVKIDFTPSQSLHGTAKDWTIVYIMDPICDTSISGLQIHLSKQNDLPV 97
DB 38 KKHIVNLCGVKIDFTPSQSLHGTAKDWTIVYIMDPICDTSISGLQIHLSKQNDLPV 97

RESULT 5 PRELIMINARY: PRT: 497 AA.
OY 076380:
DI 076380:
DB 076380:

```

AC 076380: 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, last sequence update)
DI 01-JUN-2001 (TREMblrel. 17, last annotation update)
DE ALPHA TELOMER E BINDING PROTEIN.
OS Oxytricha trifallax.
OC Eukaryota; Alveolata; Ciliophora; hypotrichs; Stichotrichida;
CC Oxytrichidae; Oxytricha.
CN NCBI_TaxID=5945;
RN [1]
RP SEQUENCE FROM N.A.
RA Prescott J.D., Dukes M.L., Prescott D.M.:
PT "Oxytricha trifallax macronuclear alpha telomere binding protein
gene".
PL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DE EMBL: AF067931; AAC7515.1; -
DB HSSP: P29549; JOTC
DR InterPro: IPR003415; Telo_bind_alpha.
DR Pfam: PF02307; Telc_bind_alpha.1;
SO SEQUENCE 497 AA: 55351 MW: 132076865C3AB566 CRC64;

Query Match 18.7% Score 105; DB 5; Length 477;
Best Local Similarity 30.4%; Pred. No. 0.0045;
Matches 31: Conservative 15; Mismatches 42; Indels 14; Gaps 2;
OY 7 TNYITFLN-OLKSGIVNYGVKFRPPYLSKGTGYCSVVIIVD-----Q 52
DB 34 SKYEVELTKADLTSVTADHFAVAVIDAFTPKINDERYICSLKIVDPYLKKEKGTG 93
OY 53 TWKLTCLFSGNFEALPIIYKNGDIVRFRLKIOYKKETO 94
DB 94 NSDYATLVYAKRFEDLPRIHRLGDIIRHRAPIRLYNGORQ 135

RESULT 6
ID 09GJ37 PRELIMINARY: PRT: 318 AA.
AC 09GJ37:
DT 01-MAY-2000 (TREMblrel. 13, Created)
DI 01-MAY-2000 (TREMblrel. 13, last sequence update)
DE 01-MAY-2000 (TREMblrel. 13, last annotation update)
GN ATG05210. PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
CC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
CN NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-GV. COLUMBIA:
RC MEDLINE=20083487; PubMed=1061197;
EX Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Icen C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.F., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.H., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., Vanaken S.F., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Cressy T.H., Goodman H.M., Sorensen C.P.,
RA Coppenhaver G.P., Preuss D., Nieman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.:
PT "Sequence and analysis of chromosome II of Arabidopsis thaliana.";
PL Nature 402:761-768(1999).
DR EMBL: AC007018; AAD29059.1; -
SO SEQUENCE 318 AA: 35220 MW: EFFIADN077B93A2E CRC64;

Query Match 15.0% Score 85; DB 10; Length 318;
Best Local Similarity 30.4%; Pred. No. 0.45;
Matches 21: Conservative 15; Mismatches 30; Indels 2; Gaps 1;
OY 23 VNYGVKFRPPYLSKGTGYCSVVIIVD--NKLTLFSGNFEALPIIYKNGDIVR 80
DB 24 VELIGVLEORPEPKPCRRNDWICILRIIDIVPSGLIVNFSKILEOLFQIKHDMIL 83

OY 81 FHLKIOYV 90
DB 84 FTRIMQIF 92

RESULT 7
ID 09Y119 PRELIMINARY: PRT: 473 AA.
AC 09Y119:
DT 01-NOV-1999 (TREMblrel. 12, Created)
DI 01-NOV-1999 (TREMblrel. 12, last sequence update)
DE 01-JUN-2001 (TREMblrel. 17, last annotation update)
GN REPLICATION PROTEIN A LARGE SUBUNIT.
RP A.
OS Cryptosporidium parvum.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
CC Cryptosporidiidae; Cryptosporidium.
CN NCBI_TaxID=5807;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=KSU-1;
RA Zhu G., Marchessa H.J., Keithly J.S.:
PT "Cryptosporidium parvum has a short-type replication protein A (RPA)
large subunit".
PL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DE EMBL: AF132307; AAD42052.1; -
DR InterPro: IPR000409; P1_PLC_X.
DR PROSITE: PS50007; P1PLC_X DOMAIN: 1.
SO SEQUENCE 473 AA: 53997 MW: DEC281A982343878 CRC64;

Query Match 13.7% Score 77.5; DB 5; Length 473;
Best Local Similarity 23.0%; Pred. No. 4.4;
Matches 26: Conservative 27; Mismatches 51; Indels 9; Gaps 4;
OY 2 SLVPAINTYITPLNOLK---GTVNYGVYKFRPP---YLSKGTGYCS--VVIIVDQI 53
DB 111 SGKAPKIVFTIVKDIKREIPVPSIVDLIGIVHFSFSTKVISRRNDEVSRRITSVDKI 170
OY 54 NWKLTCLFSGNFEALPIIYKNGDIVRFRLKIOYKKETOIGTSGFASLT 106
DB 171 GFLNITLWGLFAEVSDEKFLGNPVVALKSIQIREYQOR-OGSTLNGSNNEF 222

RESULT 8
ID 09C2K8 PRELIMINARY: PRT: 759 AA.
AC 09C2K8:
DT 01-JUN-2001 (TREMblrel. 17, Created)
DI 01-JUN-2001 (TREMblrel. 17, last sequence update)
DE 01-JUN-2001 (TREMblrel. 17, last annotation update)
GN 3H10.40.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
CC Sordariiales; Sordariaceae; Neurospora.
CN NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hohelsel J., Brandt P., Farlmann B., Holland R.,
RA Nykatura G., Maves H.W., Manhaupt G.:
PT Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
PL Nature 402:761-768(1999).
DR EMBL: AL513442; CAC28643.1; -
SO SEQUENCE 759 AA: 85218 MW: DC737A5574F4F73 CRC64;

Query Match 13.2% Score 75; DB 3; Length 759;
Best Local Similarity 25.0%; Pred. No. 14;

```

Matches 27: Conservative 24: Mismatches 41: Indels 16: Gaps 4:
OY 3 LVPATNYIYIPNL-----KGIIVGVGVKFFPKPYISKGTIDYCSVTI-----VD 51
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 18 LFPST---LIPRALLDNDGDGDPGSMVAVICVLKDCAPVAVIHSDMKCLITL.SIE 74
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 52 OTNRLTCLLFGSNGYFALFIYKNGDIVRFHRLKIOYKKKEIQTSS 99
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 75 DESAGVELVIF--PPKARPEVAGDVLVLSAKVQPKSNKSLITS 120
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

RESULT 9
O9CE7 PRELIMINARY: PRI: 383 AA.
ID O9CE7
AC O9CE7
DI 01-JUN-2001 (TRENDEL: 17, Created)
DI 01-JUN-2001 (TRENDEL: 17, Last sequence update)
DI 01-JUN-2001 (TRENDEL: 17, Last annotation update)
DE 2310015B14RIK PROTEIN.
CN 2310015B14RIK
OS Mus musculus (Mouse).
OC Eukaryota: Metazoa: Chordata: Vertebrata: Euteleostomi:
OC Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
CX NCBI_TaxID=10090:
EN 11
PP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC HEAD, AND TONGUE;
RX MEDLINE=21085650; PubMed=11217851:
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii T.,
RA Aizawa K., Iizawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito T.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochia H.,
RA Kuehl P., Lewis S., Matsumoto Y., Nikaido I., Pesole G., Osoeckem J.,
RA Schiml L.M., Stabili F., Suzuki R., Tomita M., Wagner L., Wasth G.,
RA Sakai K., Okido T., Furuno M., Aono H., Balarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
PA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
PA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee S.H.,
PA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts J.,
PA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
PA Sasaki H., Sato K., Schoenbach C., Sessa T., Shibata Y., Storch K.F.,
PA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilting L.,
PA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kchitsuki S.,
PA Hayashizaki Y.:
PI *Functional annotation of a full-length mouse cDNA collection.
PI Nature 409:585-590(2001).
PI EMBL: AK013942; BAB2027.1: -.
PI EMBL: AK003350; BAB2023.1: -.
PI MGD: MG1:1913615; 2310015B14RIK.
SO SEQUENCE 383 AA; 43208 MW; 97EF08712201E0C2 CRC64:

Query Match 13.1% Score 74.5; DB 11: Length 383;
Best Local Similarity 24.4% Pred. No. 7.2:
Matches 19: Conservative 21: Mismatches 29: Indels 9: Gaps 3:
OY 23 VVVYGVKFFPKPYISKGTIDYCSVTIIVDQIVKRL--TCLLFGSNGYFALFIYK--R-- 78
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 240 INWTVIK-----HLSFGEDIPGIVPLDHIIVTAPOASMPFOYFVAVIKVDEVLRINQSVTRHK 294
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 79 VRFHRLKIOYKKETQCI 95
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 295 LRTNPSVTRHKVANGL 312
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

RESULT 10
O9CE7 PRELIMINARY: PRI: 198 AA.
ID O9CE7
AC O9CE7
DI 01-MAY-1999 (TRENDEL: 10, Created)
DI 01-MAY-1999 (TRENDEL: 10, Last sequence update)
DI 01-MAY-1999 (TRENDEL: 10, Last sequence update)

```

```

DI 01-MAY-1999 (TRENDEL: 10, Last annotation update)
DE HYPOTHEICAL 22.4 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
OC Mammalia: Eutheria: Primates: Catarrhini: Hominiidae: Homo.
CX NCBI_TaxID=9606:
EN 11
PP SEQUENCE FROM N.A.
RA Barrow J.K.-P., Boguski M.S., Touchman J., Spencer F.,
PI *Full-insert sequence of mapped XREF EST.
PI Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DE EMBL: AF091085; AAC72954.1: -.
KW Hypothetical protein
SO SEQUENCE 198 AA; 22396 MW; 75CE7A231C91EF02 CRC64:

Query Match 13.1% Score 74; DB 4: Length 198;
Best Local Similarity 26.2% Pred. No. 3.7:
Matches 17: Conservative 18: Mismatches 26: Indels 4: Gaps 2:
OY 36 YISKGTIDYCSVTIIVDQIVKRL--TCLLFGSNGYFALFIYK--NGDIVRFHRLKIOYKK 91
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 111 HLSFGEDIPGIVPLDHIIVTAPOASMPFOYFVAVIKVDEVLRINQSVTRHK 170
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 32 ETQCI 95
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 171 VANGL 175
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

RESULT 11
O9CE7 PRELIMINARY: PRI: 228 AA.
ID O9CE7
AC O9CE7
DI 01-JUN-2001 (TRENDEL: 17, Created)
DI 01-JUN-2001 (TRENDEL: 17, Last sequence update)
DI 01-JUN-2001 (TRENDEL: 17, Last annotation update)
DE D347704.2.2 (CGI-54 (LOC51614), ISOFORM 2) (FRAGMENT).
GN D347704.2.
OS Homo sapiens (Human).
OC Eukaryota: Metazoa: Chordata: Vertebrata: Euteleostomi:
OC Mammalia: Eutheria: Primates: Catarrhini: Hominiidae: Homo.
CX NCBI_TaxID=9606:
EN 11
PP SEQUENCE FROM N.A.
RA Brown A.:
PI Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DE EMBL: AL121586; FACC3497.1: -.
PI NON-ITER 1
SO SEQUENCE 228 AA; 25918 MW; 97888251C9A7B93 CRC64:

Query Match 13.1% Score 74; DB 4: Length 228;
Best Local Similarity 26.2% Pred. No. 4.4:
Matches 17: Conservative 18: Mismatches 26: Indels 4: Gaps 2:
OY 36 YISKGTIDYCSVTIIVDQIVKRL--TCLLFGSNGYFALFIYK--NGDIVRFHRLKIOYKK 91
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 139 HLSFGEDIPGIVPLDHIIVTAPOASMPFOYFVAVIKVDEVLRINQSVTRHK 198
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 52 ETQCI 95
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 199 VANGL 203
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

RESULT 12
O9CE7 PRELIMINARY: PRI: 346 AA.
ID O9CE7
AC O9CE7
DI 01-JUN-2000 (TRENDEL: 15, Created)
DI 01-JUN-2000 (TRENDEL: 15, Last sequence update)
DI 01-JUN-2000 (TRENDEL: 15, Last annotation update)

```

```

OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
OC Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
PC TISSUE=LIVER;
RA Zhang G., Yu Y., Zhang S., Wei H., Zhou G., Ouyang S., Luo L., Bi J.,
RA Liu M., He F.;
RT *Functional prediction of the coding sequences of 121 new genes
RT deduced by analysis of cDNA clones from human fetal liver.;
PL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF115614; AAF1038.1;
SO SEQUENCE 346 AA: 39238 MW: A042221C998834BF CRC64;

Query Match 13.1%; Score 74; DB 4; Length 345;
Best Local Similarity 26.2%; Pred. No. 7.2;
Matches 17; Conservative 18; Mismatches 26; Indels 4; Gaps 2;

OY 36 YLSKGDICSVTVIVDQINVKL--TCLLFGNVEALPIIK--NGDIVRFHRLKIVYKK 91
DB 211 HLSFGEDYPGIVPLDHTNIVAPQASMMFYFVYKVPVIVYKVDGVALRTNQSIVIPHEK 270
OY 92 ETGCI 96
DB 271 VANGI 275

RESULT 13
OY282
ID 09Y282 PRELIMINARY: PRI: 383 AA.
AC 09Y282.
DI 01-NOV-1999 (IREMBLrel. 12, Created)
DI 01-NOV-1999 (IREMBLrel. 12, Last sequence update)
DI 01-MAR-2001 (IREMBLrel. 16, Last annotation update)
DE CGI-54 PROTEIN (DJ47704.2) (CGI-54).
GN DJ47704.2.
OS Homo sapiens (Human).
OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
OC Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lin W.-C.;
RT *Comparative gene cloning: Identification of novel human genes with
RT Caenorhabditis elegans proteome as template.*;
PL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Zhang J., Liu T., Ye M., Zhang Q., Fu G., Zhou J., Wu J., Shen Y.,
RA Yu M., Chen S., Mao M., Chen Z.;
RT *Human hypothetical 43.2 kd protein.*;
PL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Brown A.;
PL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF151812; AAD34049.1;
DR EMBL: AF077030; AAD27763.1;
DR EMBL: AL121586; CAB89412.1;
SO SEQUENCE 383 AA: 43222 MW: 8370FB3067AD95CE CRC64;

Query Match 13.1%; Score 74; DB 4; Length 393;
Best Local Similarity 26.2%; Pred. No. 8.1;
Matches 17; Conservative 18; Mismatches 26; Indels 4; Gaps 2;

OY 36 YLSKGDICSVTVIVDQINVKL--TCLLFGNVEALPIIK--NGDIVRFHRLKIVYKK 91
DB 248 HLSFGEDYPGIVPLDHTNIVAPQASMMFYFVYKVPVIVYKVDGVALRTNQSIVIPHEK 307
OY 92 ETGCI 96

```

```

DB 308 VANGI 312

RESULT 14
OY281
ID 09Y281 PRELIMINARY: PRI: 299 AA.
AC 09Y281.
DI 01-OCT-2000 (IREMBLrel. 15, Created)
DI 01-OCT-2000 (IREMBLrel. 15, Last sequence update)
DI 01-OCT-2000 (IREMBLrel. 15, Last annotation update)
DE BACM4124.C.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota: Metazoa: Arthropoda: Insecta:
OC Ephyroidea: Neuroptera: Endopterygota: Diptera: Brachycera: Muscomorpha:
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Møller, Gadieu, Dreano, Lelaure, Galibert F.;
RT *Sequencing the distal X chromosome of Drosophila melanogaster.*;
PL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Sence F.;
PL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL121805; CAB58072.1;
SO SEQUENCE 299 AA: 32007 MW: C7224AFCDC0809B6 CRC64;

Query Match 12.8%; Score 72.5; DB 5; Length 299;
Best Local Similarity 33.3%; Pred. No. 8.8;
Matches 19; Conservative 14; Mismatches 19; Indels 5; Gaps 1;

OY 52 GIVNKTICLIFSGNVEALPIIK-----NGDIVRFHRLKIVYKKTCGTCSTGAS 103
DB 21 KNSLLTLLHLLHLENSKIVKPKPLGNSNITIKKIKIKVKTSTMSSTGVAS 77

RESULT 15
OY289
ID 09Y289 PRELIMINARY: PRI: 937 AA.
AC 09Y289.
DI 01-OCT-2000 (IREMBLrel. 15, Created)
DI 01-OCT-2000 (IREMBLrel. 15, Last sequence update)
DI 01-JUN-2001 (IREMBLrel. 17, Last annotation update)
DE F22G5.6.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae:
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shing P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
RA Kim G., Altaji H., Bel O., Ching C., Chiu J., Choi E., Conn L.,
RA Conner A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
RA Leng C., Li J., Liu A., Liu K., Liu S., Mukhariskiy N., Nguyen M.,
RA Paim C., Phan P., Sakano H., Schwartz J., Southwick A., Thayer A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RT *Genetic sequence for Arabidopsis thaliana BAC F22G5 from chromosome
RT 1.*;
PL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
PL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
PL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 12, 2002, 08:33:16 ; Search time 12.47 Seconds

(without alignments)
195.701 Million cell updates/sec

Title: US-09-816-248-5

Perfect score: 567

Sequence: 1 MSLEPAIMYITPLNOLKGC.....KKEIGITSSGSLIFEET 109

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2090000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_5/ptodata/2/1aa/5A.COMB.pep:*
- 2: /cgn2_5/ptodata/2/1aa/5B.COMB.pep:*
- 3: /cgn2_5/ptodata/2/1aa/5A.COMB.pep:*
- 4: /cgn2_5/ptodata/2/1aa/5B.COMB.pep:*
- 5: /cgn2_5/ptodata/2/1aa/PTUS.COMB.pep:*
- 6: /cgn2_5/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	70.5	12.4	287	3	US-08-733-230-2
2	70.5	12.4	287	4	US-08-953-326-2
3	64.5	11.4	412	2	US-08-741-134-2
4	62	10.9	990	2	US-08-645-153B-15
5	61.5	10.8	155	2	US-08-162-402B-18
6	60	10.6	990	2	US-08-392-625-20
7	60	10.6	990	2	US-08-466-951A-20
8	59	10.4	405	1	US-07-996-772A-4
9	59	10.4	405	1	US-08-445-822-4
10	59	10.4	405	1	US-08-480-229C-5
11	58	10.2	85	2	US-08-659-235C-5
12	58	10.2	1169	1	US-08-642-921-2
13	58	10.2	1169	2	US-08-880-685-2
14	58	10.2	1169	2	US-08-880-684-2
15	58	10.2	1169	2	US-08-152-721B-2
16	57.5	10.1	383	2	US-08-569-168-7
17	57	10.1	387	1	US-07-996-772A-2
18	57	10.1	387	1	US-08-446-822-2
19	57	10.1	387	5	PCT-US93-12586-2
20	57	10.1	387	5	PCT-US93-915-2
21	56.5	10.0	504	1	US-07-932-915-2
22	56.5	10.0	504	1	PCT-US91-05825-2
23	56.5	10.0	542	4	US-09-189-462-6
24	56.5	10.0	906	1	US-08-220-151-9
25	56.5	10.0	906	1	US-08-413-118-9
26	56.5	10.0	906	1	US-08-473-446-2
27	56.5	10.0	907	1	US-08-349-006-2

28	56.5	10.0	907	3	US-08-804-439A-19	Sequence 19, App1
29	56.5	10.0	907	3	US-08-720-229-19	Sequence 19, App1
30	56.5	10.0	907	5	PCT-US94-04180-2	Sequence 21, App1
31	56	9.9	331	2	US-08-878-989-21	Sequence 21, App1
32	56	9.9	331	3	US-09-101-146-64	Sequence 64, App1
33	56	9.9	331	4	US-09-272-795-21	Sequence 21, App1
34	56	9.9	331	4	US-07-783-705A-4	Sequence 4, App1
35	55.5	9.8	863	4	US-09-238-303-11	Sequence 11, App1
36	55.5	9.8	592	3	US-08-991-813-2	Sequence 2, App1
37	54.5	9.6	213	2	US-08-716-284-2	Sequence 2, App1
38	54.5	9.6	251	1	US-08-425-336-100	Sequence 100, App
39	54.5	9.6	251	1	US-08-488-113B-100	Sequence 101, App
40	54.5	9.6	251	1	US-08-477-484B-100	Sequence 100, App
41	54.5	9.6	251	1	US-08-477-484B-101	Sequence 101, App
42	54.5	9.6	251	2	US-08-646-360-100	Sequence 100, App
43	54.5	9.6	251	2	US-08-646-360-101	Sequence 101, App
44	54.5	9.6	251	2	US-08-646-360-101	Sequence 101, App
45	54.5	9.6	251	2	US-08-646-360-101	Sequence 101, App

ALIGNMENTS

RESULT 1
US-08-733-230-2
: Sequence 2, Application US/08733230
: Patent No. 6025338
GENERAL INFORMATION:
: APPLICANT: Barbey, Anthony F.
: APPLICANT: Ganla, Roman Reddy
: APPLICANT: McGuire, Travis C.
: APPLICANT: Burridge, Michael J.
: APPLICANT: Nylka, Aceme
: APPLICANT: Rurangirwa, Fred R.
: TITLE OF INVENTION: Nucleic Acid Vaccines Against
: TITLE OF INVENTION: Rickettsial Diseases and Methods of Use
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Salivanchik & Salivanchik
: STREET: 2421 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: USA
: ZIP: 32606
COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/733,230
: FILING DATE:
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Whitlock, Ted W.
: REGISTRATION NUMBER: 36,965
: REFERENCE/Docket NUMBER: UF-167
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 352-375-8100
: TELEFAX: 352-372-5800
: INFORMATION FOR SEQ. ID NO.: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 287 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-733-230-2

Query: Match 12.4% Score 70.5; DB 3; Length 287;
Best Local Similarity 30.3%; Pred. No. 0.47;
Matches 27; Conservative 19; Mismatches 18; Indels 25; Gaps 7;

```

OY      35 PYLTK--GIDTCSVTIVDOINQTLIC-----LFLESCNEALPII::KRGDIYRPHPL---84
          ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db     194 PYCAGIGTD--LVSVINAIIPKLSTVQKGKLGISYSINSLA--SIITGS---HFRIRVIGH 246
          ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
OY      85 -----KIIOVKREFOGIITSSGFASFLFE 107
          .....:|||||:|||||:
Db     246 EFKDIIATLIIFTSKI-GISNPGFASATLD 273

```

RESULT 2
US-08-953-326-2

```

Sequence 2 Application US/08953326
Patent No. 6251872
GENERAL INFORMATION:
APPLICANT: Barbel, Anthony F.
APPLICANT: Ganta, Roman P.
APPLICANT: McGuire, Travis C.
APPLICANT: Burridge, Michael J.
APPLICANT: Myka, Aceme
APPLICANT: Rurangirwa, Fred R.
APPLICANT: Mahan, Susan M.
TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
TITLE OF INVENTION: Animals and Humans
FILE REFERENCE: UF-157C1
CURRENT APPLICATION NUMBER: US/08/953,326
CURRENT FILING DATE: 1997-10-17
EARLIER APPLICATION NUMBER: 08/953,326
EARLIER FILING DATE: 1997-10-17
EARLIER APPLICATION NUMBER: 08/733,230
EARLIER FILING DATE: 1995-10-17
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 2
LENGTH: 287
TYPE: PPT
ORGANISM: Coxiella burnetium
US-08-953-326-2

```

```
Query Match 12.4% Score 70.5; DB 4 Length 287;  
Best Local Similarity 30.3%; Pred. No. 0.67;  
Matches 27: Conservative 19; Mismatches 18; Indels 25; Gaps 7  
  
OY      35 PYLTK--GDYSVCVIVDQINAKTC-----LLFSGREALPIIYKRQDVPHRL--- 84  
          || :||:::||| | | | | | | | |  
Db    194 PLYCAGSGLGID---LSVNIAHPKSLSTOGKGISINSSEA-SIFIGS--HPRVIGN 245  
  
OY      85 -----KIQQYYKETGGTSGFASLTTE 107  
          ::::: | | | | | | | |  
Db   246 ERFDIATLKLFITSKT-GISENGPASFATIID 273
```

RESULT 3
 US-08-741-134-2
 : Sequence 2, Application US/08741134
 : Patent No. 5861498
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Litack, Gerald
 : APPLICANT: Altemir, Enad S.
 : APPLICANT: Fernandez-Almoncri, Teresa
 : TITLE OF INVENTION: IMMUNOPHILIN FRBP45 AND COMPOSITIONS FOR MAKING
 : TITLE OF INVENTION: AND
 : TITLE OF INVENTION: METHODS OF USING THE SAME
 :
 : NUMBER OF SEQUENCES: 6
 :
 : CORRESPONDENCE ADDRESS:
 : ADDRESS: Woodcock Washburn Kurtz Mackiewicz & No. 5861498r1s
 : STREET: One Liberty Place - 45th floor
 : CITY: Philadelphia
 : STATE: Pennsylvania
 : COUNTRY: USA
 :
 : ZIP: 19103
 :
 : COMPUTER READABLE FORM:

```

: MEDIUM TYPE: floppy disk
: COMPUTER: IBM pc compatible
: OPERATING SYSTEM: Windows 3.11
: SOFTWARE: WordPerfect for Windows 6.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/741.134
: FILING DATE:
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/007,163
: FILING DATE: 01-NOV-1995
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Deluca, Mark
: REGISTRATION NUMBER: 33,229
: REFERENCE/DOCKET NUMBER: JUJ-2090
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 215-568-3100
: TELEFAX: 215-568-3439
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 412 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-741.134-2

```

```

Query Match: 11.43: Score 64.5; DB 2: Length 412:
Best Local Similarity 32.2%; P-Val 4.5e-06;
Matches 12: Conservative 9; Mismatches 20; Indels 11; Gaps 3
OY 33 KPFPLNG-----TDYCVYTIYDITNKKITGLFSGNYELP--IIYKGDIVRF 81
19 KPFHISALANDISTGNDPQVYVYVYDKNF-LVCTLOKGI IGVPLDLFLKSGDSVF 75
Db

```

RESULT 4
US-08-545-1938-15
Sequence 15, Application US/08645193B
Patent No. 5965253
GENERAL INFORMATION:
APPLICANT: Kupke, Thomas
APPLICANT: Golz, Friedrich
APPLICANT: Kempler, Christoph
APPLICANT: Jung, Gunter
TITLE OF INVENTION: Oxidative Decarboxylation of Peptides
TITLE OF INVENTION: Catalyzed by Flavoprotein Epub
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Steiner, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
Zip: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/545,193B
FILING DATE: 13-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 0652.1540000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 15:

Db 794 -KMKKDYCLNSELVDYSIVDYVPEVRYGCPHYIEDIENF 833

RESULT 7

US-08-466-961A-20
: Sequence 20, Application US/08466561A
: Patent No. 5843709
: GENERAL INFORMATION:
: APPLICANT: Entian, Karl-Dieter
: APPLICANT: Gtz, Friedrich
: APPLICANT: Schnell, No. 5843709bert
: APPLICANT: Augustin, Johannes
: APPLICANT: Engelke, Gernar
: APPLICANT: Rosenstein, Ralf
: APPLICANT: Kaletta, Corina
: APPLICANT: Klein, Cora
: APPLICANT: Wieland, Bernd
: APPLICANT: Kupke, Thomas
: APPLICANT: Jung, G nther
: APPLICANT: Kellner, Poland
: TITLE OF INVENTION: Biosynthetic Process for the Preparation of
: TITLE OF INVENTION: Chemical Compounds
: NUMBER OF SEQUENCES: 42
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sterne, Kessler, Goldstein & Fox
: STREET: 1100 New York Avenue, NW
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.A.
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/466,961A
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/392,625
: FILING DATE: 22-FEB-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/876,791
: FILING DATE: 30-APR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/784,234
: FILING DATE: 31-OCT-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Esmond, Robert W.
: REGISTRATION NUMBER: 32,893
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 371-2600
: TELEFAX: (202) 371-2540
: INFORMATION FOR SEQ ID NO: 20:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 990 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: US-08-466-961A-20

Query Match 10.5% Score 60: DB 2: Length 990:
Best Local Similarity 24.5% Pred. No. 59:
Matches 25: Conservative 16: Mismatches 35: Indels 25: Gaps 5:

0Y 4 VPAI-----NYITPLNLSKGTIVNYGVK-----FFKPYLSGTDYCSVTIV 60
Db 734 IPTIYDNIYDILLPLFIILKYNNEINFFYIKREDEDFILRLPELEIDYQSIVSFI 733
0Y 51 DQINVLKLLFSGRY-----EALPIIKNG-----DIVPF 81

Db 794 -KMKKDYCLNSELVDYSIVDYVPEVRYGCPHYIEDIENF 833

RESULT 8

US-07-996-772A-4
: Sequence 4, Application US/07996772A
: Patent No. 5472866
: GENERAL INFORMATION:
: APPLICANT: Gerald, Christophe
: APPLICANT: Hartig, Paul R.
: APPLICANT: Branchek, Theresa A.
: APPLICANT: Weinshank, Richard L.
: TITLE OF INVENTION: DNA ENCODING 5-HT4a SEROTONIN
: TITLE OF INVENTION: RECEPTORS AND USES THEREOF
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: COOPER & DUNHAM
: STREET: 30 ROCKEFELLER PLAZA
: CITY: NEW YORK
: STATE: NEW YORK
: ZIP: 10112
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.24
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/996,772A
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: White, P. John
: REGISTRATION NUMBER: 28,678
: REFERENCE/DOCKET NUMBER: 42667/JPW/TEP
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 977-9550
: TELEFAX: (212) 664-0525
: TELEX: 422523 COOP U1
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 406 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-07-996-772A-4

Query Match 10.4% Score 59: DB 1: Length 406:
Best Local Similarity 28.4% Pred. No. 23:
Matches 23: Conservative 12: Mismatches 30: Indels 16: Gaps 5:

0Y 7 TWITITPL--NOLKGTIVNYGVKFFKPYLSKGTIDYCSVTIVDQINVKLT----- 58
Db 55 TWFTIVSLAFADLLVSYLVNAFMAELIVODIWF-YGEMFCIVRISLD--VLLITISIFH 110
0Y 59 -CLIFSGNTEAL---PIIKYN 75
Db 111 LCCLSLDRIVAIICOPIVYRN 131

US-08-446-822-4
: Sequence 4, Application US/08446822
: Patent No. 5766879

: GENERAL INFORMATION:
: APPLICANT: SYNAPTIC PHARMACEUTICAL CORPORATION
: TITLE OF INVENTION: DNA ENCODING 5-HT4 SEROTONIN RECEPTORS
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: COOPER & DUNHAM
: STREET: 30 ROCKEFELLER PLAZA

CITY: NEW YORK
STATE: NEW YORK
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,822
FILING DATE: June 1, 1995
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: White, P., John
REGISTRATION NUMBER: 28,578
REFERENCE/DOCKET NUMBER: 42667-A-PCT-US/JPW/MAT
TELEPHONE: (212) 278-0400
TELEFAX: (212) 291-0525
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 406 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-446-822-4

Query Match 10.4% Score 59; DB 1; Length 406;
Best Local Similarity 28.4%; Pred. No. 23;
Matches 23; Conservative 12; Mismatches 30; Indels 6; Gaps 5;

OY 7 TNYITPL--NOLKGTIVNVGVVFFKPPYLSKGTIDYCSVTIVDTNWKLT----- 56
DB 55 TNFIVSLAFADLVSLVNAFGAIELVODIMF--YGEMLVLRISID---VLTITASTIFH 110
OY 59 -CLFSGNEAL--PIIYKN 75
DB 111 LCCLSLDRIYVAICCPLYVRN 131

RESULT 10
PCT-US93-12586-4
Sequence 4, Application PC/TUS9312586
GENERAL INFORMATION:
APPLICANT: SYNAPIC PHARMACEUTICAL CORPORATION
TITLE OF INVENTION: DNA ENCODING 5-HT4 SEROTONIN RECEPTORS
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESSES:
ADDRESSEE: COOPER & DUNHAM
STREET: 30 ROCKEFELLER PLAZA
CITY: NEW YORK
STATE: NEW YORK
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/12586
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: White, P., John
REGISTRATION NUMBER: 28,578
REFERENCE/DOCKET NUMBER: 42667-A-PCT-US/JPW/IEP
TELEPHONE: (212) 977-9550
TELEFAX: (212) 564-0525
TELEX: 422523 COOP UT
INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 406 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-12586-4

Query Match 10.4% Score 59; DB 5; Length 406;
Best Local Similarity 28.4%; Pred. No. 23;
Matches 23; Conservative 12; Mismatches 30; Indels 16; Gaps 5;

OY 7 TNYITPL--NOLKGTIVNVGVVFFKPPYLSKGTIDYCSVTIVDTNWKLT----- 58
DB 55 TNFIVSLAFADLVSLVNAFGAIELVODIMF--YGEMLVLRISID---VLTITASTIFH 110
OY 59 -CLFSGNEAL--PIIYKN 75
DB 111 LCCLSLDRIYVAICCPLYVRN 131

RESULT 11
US-08-480-229C-5
Sequence 5, Application US/08480229C
Patent No. 5874562
GENERAL INFORMATION:
APPLICANT: Quentromus, Thomas
APPLICANT: Hogan, Brigid
APPLICANT: Snodgrass, H. Ralph
APPLICANT: Zupancic, Thomas J.
TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
TITLE OF INVENTION: CELL LOCUS-1
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,229C
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 8907-0026-999
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 Pennie
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 85 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-480-229C-5

Query Match 10.2% Score 58; DB 2; Length 85;
Best Local Similarity 30.0%; Pred. No. 3; 4;
Matches 18; Conservative 10; Mismatches 24; Indels 8; Gaps 2;

OY 29 VYFKPPYISGTYGCSVTIVDTNWKLTLLFSGNEALPIIYK---NDIVRFRRLK 85
DB 111 LCCLSLDRIYVAICCPLYVRN 131

Db 27 VKSKYKDSSNGEDMIXILKGNKH-----LVFTGHTDAIDVYRPFSPVITTFVRLR 81

RESULT 12

US-08-659-235C-5
Sequence 5, Application US/08659235C
Patent No. 5877281

GENERAL INFORMATION:

APPLICANT: Quartermou, Thomas
APPLICANT: Hogan, Bridgid
APPLICANT: Snodgrass, H. Ralph
APPLICANT: Zupancic, Thomas J.
TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
TITLE OF INVENTION: CELL LOCUS-1
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/659,235C
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Poissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 8907-0034-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 859-8854/9741
TELEX: 66141 Pennie

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:
LENGTH: 85 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown

MOLECULE TYPE: peptide
US-09-659-235C-5

Query Match 10.2% Score 58; DB 2; Length 85;
Best Local Similarity 30.0%; Pred. No. 3.4;

Matches 18: Conservative 10: Mismatches 24: Indels 8: Gaps 2:

Db 27 VKSKYKDSSNGEDMIXILKGNKH-----LVFTGHTDAIDVYRPFSPVITTFVRLR 81

RESULT 13

US-08-542-921-2
Sequence 2, Application US/08542921
Patent No. 5736514

GENERAL INFORMATION:

APPLICANT: IIZUKA, TOSHIHIKO
APPLICANT: TAGAWA, MICHITO
APPLICANT: ARAI, SATOSHI
APPLICANT: MIYAKE, MASATSUGU
APPLICANT: MIYAKE, TOSHIRO

TITLE OF INVENTION: NOVEL BACILLUS STRAIN AND HARBOR

TITLE OF INVENTION: ORGANISM CONTROLLING AGENTS

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT

STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/542,921
FILING DATE: 13-OCT-1995
CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 276082/94
FILING DATE: 14-OCT-1994
ATTORNEY/AGENT INFORMATION:

NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 49-209-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 24855 OPAT UR

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 119 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-542-921-2

Query Match 10.2% Score 58; DB 1; Length 119;
Best Local Similarity 32.4%; Pred. No. 1.3e+02;

Matches 23: Conservative 9: Mismatches 31: Indels 8: Gaps 4:

Db 35 YLSKGTIDYGVVITVDIOTWKLTLFSGNYEALPIYKNGDIVR-FHRLK-IGVYKKE 92

Db 439 YLSSG--QVSGISGYTQCIIPAVCLQORNSDLPINFGEDJINNSHRLSHITQYRFQ 455

Db 93 TCGTSSGFAS 103

Db 497 A---TQSGSPS 504

RESULT 14

US-08-690-685-2
Sequence 2, Application US/08880685
Patent No. 5834296

GENERAL INFORMATION:

APPLICANT: IIZUKA, TOSHIHIKO
APPLICANT: TAGAWA, MICHITO
APPLICANT: ARAI, SATOSHI
APPLICANT: MIYAKE, MASATSUGU
APPLICANT: MIYAKE, TOSHIRO

TITLE OF INVENTION: NOVEL BACILLUS STRAIN AND HARBOR

TITLE OF INVENTION: ORGANISM CONTROLLING AGENTS

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

Email: coapbs-r@mail.nih.gov
 Tissue Procurement: ATCC/DCTD/DTP
 cDNA Library Preparation: Rubin Laboratory
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc_mgc@hgrl.nih.gov
 Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
 Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
 Lim, M., Maduro, O.L., Mastello, C., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J.J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,
 Tjongson, E.E., Touchman, J.W., Tsurgon, C., Vogt, J.L., Walker, M.A.,
 Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 15 Row: 9 Column: 18
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 10434050.

FEATURES

Location/Qualifiers

1..2029
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="MGC:10280 IMAGE:3955573"
 /tissue_type="Skin, melanotic melanoma."
 /clone_lib="NIH-MGC-20"
 /lab_host="DH10B-R"
 /note="Vector: pOTB7"
 67..1971
 /codon_start=1
 /product="Unknown (protein for MGC:10280)."
 /protein_id="AAH02923.1"
 /db_xref="GI:12804139"
 /translation="MSLVPAIYIYPLNQLKGGIVNYGVVKKFPYLSKGDYD
 SVYIVDQNTALICLLFSNYEALPIYKNGDIVRHLKIOVKKETOITSSGFA
 SLTFEGTGLVPIIPRTSKYFNFTEDHKMVEALRWASTHSPSWTLKLCGVPMQ
 YDLTCLLGAKEAVDGSFLKAVWDGTRTPPFSRVLLDVLVLEGLDLSHRLQNTI
 DILVDNHVARSLLKVSFLRIYLSLTKLOSNSENOTMLSEFLHGGTGYGRIR
 VLPESNDVDQLKLESANLTANQHSVICOSEPDDSPSSGSVSLYEVECOQLSA
 TILTDHYELRPLCAILKQAPQYRIRAKRSYKPRFLQSVKHLKPKCHLLOEVP
 HEGDDLIIFDQDATTIDYKQNTSLYDSKIWTRNOKRKVAVHFVKNGLILPSNE
 CLLLEGGTSLSEICKLSNKFNSVIVPSGHEDLELLDLSAPFLIGTTHYCKDCSS
 LRSIOLNLSVDTKSWIPSSVAEALGIVPLQYFVMTFLDDGTCGLVLEAYLMDSKFF
 QIPASFLMDDDLQKSDVMDIMDFCPGPKIDAYPWLCEFIKSYNVTNGTDNQICIQI
 FDTTVAEDVI"
 655 a 391 c 386 g 597 t

BASE COUNT

ORIGIN

Query Match 80.7%; Score 1048; DB 9; Length 2029;
 Best Local Similarity 90.8%; Pred. No. 7.5e-268;
 Matches 1178; Conservative 0; Mismatches 0; Indels 120; Gaps 1;
 1 atgtctttggtccagcaaaattatatatcacaccctgaatcaacttaagggtgt 60
 67 ATGCTTTGGTTCACCAAAATATATATATACACCCTGAATCAACTTAAGGTGT 126
 QY 61 acaattgtcaatgtctatgtgtgtgaagttctttaagcccccatatcaagga 120
 DB 127 ACAATTTGTCATGTCATGGTGTGTGAAGTCTTTTAAGCCCCCATATATCAAGCAAGGA 186
 QY 121 actgattattgtcagttgaattgttgagaccagacaaatgtataaactaactgcctg 180
 DB 187 ACTGATTATTGCTCAGTGTGAATATTGTGGACACACAAATGTAAACATACTTGCCCTG 246
 QY 181 ctcttttagtggaaactatgaagcccttccaataattataaaatggagattgttctgc 240
 DB 247 CTCTTTAGTGGAAACTATGAAGCCCTTCCAATAAATTTATAAAATGAGATATTGTTCOC 306
 QY 241 ttccacaggctgaagattccaagtataaaaggagactcagggtatccaccagctctggc 300

RESULT 2

Db 307 TTTACAGGCTGAAGATTCAAGTATATAAAAGGAGACTCAGGGTATACACGCTCTGGC 366
 QY 301 ttgcatctttgacgttttgaggaacttttgaggagccctatcatcactcgcactcaagc 360
 Db 367 TTTGCACTTTTACGTTTGAGGGAACCTTTGGGAGCCCTATCATACCTCCACTTCAAGC 426
 QY 361 aagtattttaacttactactgagggaccacaaatggtagaagccttacgtgtttgggca 420
 Db 427 AAGTATTTTAACTTCACTACTGAGGACCACAAATGGTAGAACCTTACGTGTTTGGGCA 486
 QY 421 tctactcatagtcaccgctttgagcattactactaaaaattgtgtgatgttcagccaatgcag 480
 Db 487 ICTACTCATATGTCACCGCTCTTGACATTACTATAAATTTGTGTATGTTTACGCAATGCAG 546
 QY 481 tatttgacctgactgtcagctcttgaggcaagcagaagtgacgagcattctttt 540
 Db 547 TATTTTGACCTGACTTGTCACTCTTTGGGCAAGCAGAGTGGACGGAGCATCATTTCTT 606
 QY 541 ctaaggatggatggcaccaggacaccatttcccatcttggagagctcttaataacaagac 600
 Db 607 CTAAGGTATGGATGGCACCAGCACACCATTTCCATCTTTGGAGAGTCTTAATACAGAC 666
 QY 601 ctgttcttgaagtgattttaagtcacatccatcggtcacaaaaatctgacaatagacatt 660
 Db 667 CTGTGTTCTTGAAGGTGATTTAAGTCACATCCATCGGCTACAAAATCTGACAAATAGACAT 726
 QY 661 ttagtctacgataaacatgttcatgtggcaagatctctgaagttggaagcttcttaaga 720
 Db 727 TTAGTCTACGATAAACCATGTTTCATGTGGCAAGATCTGGAAGTTGGAAGCTTCTTAGA 786
 QY 721 atctatagcttcatcaccacaaacttcaatgaattcagagagatcagacaatgtttaagt 780
 Db 787 ATCTATAGGCTTTCATACCAAACTTCAATCAATGAATTCAGAGAATCAGACAATGTTAAGT 846
 QY 781 tttaggtttcatcttcatggaggtaccagttacggtcggggaatcaggtcttgcagaa 840
 Db 847 TTAGAGTTTCATCTTCATGGAGGTACAGTTACGGTCGGGGAATCAGGCTCTTGCAGAA 906
 QY 841 agtaactctgatgtggtatcaactgaaaaaggatttagaactctgaaaatttgacagccaat 900
 Db 907 AGTAACCTCTGATGTGATCAACTCACTGAAAAGGATTTAGAAATCTGCAAAATTTGACAGCCAAT 966
 QY 901 cagcattcagatgttatctgtcaatcagaacctgacagcagctttccaaaatggagctctcg 960
 Db 967 CAGCATTCAGATGTTATCTGTCAATCAGAACCTGACGACAGCTTTTCCAA----- 1015
 QY 961 ctctgctctccaggtctggagttcagtggcaggtctctcggtcatctgcagctccacctcc 1020
 Db 1016 ----- 1015
 QY 1021 tgagttcaagctctctctgctcctcagcctcccaagtagctgggttacaggctctggatca 1080
 Db 1016 -----GCTCTGGATCA 1026
 QY 1081 gtatcattatcagaggtagaagaagtcaacagctatctgtacaatacttacaagatcat 1140
 Db 1027 GTATCATTTACAGGTAGAAAGATGTCAACAGCTATCTGCTACAAATCTTACAGATCAT 1086
 QY 1141 cagttattggagaggacaccactatgtgccatttggaaacaaaaggctcctcaacaatc 1200
 Db 1087 CAGTATTTGGAGAGGACACCACCTATGTGCCATTTTGAACAAAAGCTCTCAACAATAC 1146
 QY 1201 cgcattccgagcaaaattgaggttcataaagccccagagaactatttcagctctgttaactt 1260
 Db 1147 CCATCCGAGCAAAATGAGGTTCATATAGCCCAAGACTATTTCAGTCTGTAAACTT 1206
 QY 1261 cattgccctaaatgtcatttctgcaagaagtctccaca 1298
 Db 1207 CATTCGCCCTAAATGTCATTGCTGCAAGAGTTCACACA 1244

[illegible][illegible]

oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

XX
SQ Sequence 866 BP; 254 A; 162 C; 171 G; 274 T; 5 other;

Query Match 26.3%; Score 342; DB 22; Length 866;
Best Local Similarity 91.2%; Pred. No. 3.9e-93;
Matches 384; Conservative 0; Mismatches 30; Indels 7; Gaps 2;

QY 1 atgtcttgggtccagcaacaattatatatatatacacccctgaatcaacttaagggtggt 60
DB 439 atgtcttgggtccagcaacaattatatatatatacacccctgaatcaacttaagggtggt 498
QY 61 acaattgtcaatgtctatgtgtgtgaagttctttaagcccccatacttaagcaaga 120
DB 499 acaattgtcaatgtctatgtgtgtgaagttctttaagcccccatacttaagcaaga 558
QY 121 actgattattgtcagttgaactattgtggaccagacaaatgttaaaactaacttgctcgt 180
DB 559 actgattattgtcagttgaactattgtggaccagacaaatgttaaaactaacttgctcgt 618
QY 181 ctcttaagggaacatgaagcccttcataattataaaatgagatattgttcgc 240
DB 619 ctcttaagggaacatgaagcccttcataattataaaatgagatattgttcgc 678
QY 241 ttccacaggctgaagattcaagtataaaaggagacacagggtatccaccagctctg 300
DB 679 ttccacaggctgaagattcaagntntaaaggagacacagggtatccaccagctctg 738
QY 301 ttgcattcttgacgtttgaggaaactttggagagcccttatcatactgcactcaagc 360
DB 739 ttgcattcttgacgtt--gagggaactttggagagcccttatcatactgcactnaagcc 796
QY 361 aagtatttaacttcaactctgagggaccacaaatgttagaagcctttagtggtgggca 420
DB 797 agtatttacttctact----gaggccncaaatggnagagccttacgggttggtggt 851
QY 421 t 421
DB 852 t 852

RESULT 8
AA161067
ID AA161067 standard; cDNA; 2778 BP.
AC AA161067;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 5056.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; anyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukaemia; SS.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.

XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-052317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR P-PSDB; AAH41911.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
PS Claim 1; SEQ ID NO 5056; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAH38642-AAH42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 2778 BP; 876 A; 495 C; 487 G; 920 T; 0 other;

Query Match 15.3%; Score 199; DB 22; Length 2778;
Best Local Similarity 100.0%; Pred. No. 1.1e-49;
Matches 199; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1100 aaagatgtcaacagctatctgtacataactacagatcatcagttatggagagacac 1159
DB 1 aaagatgtcaacagctatctgtacataactacagatcatcagttatggagagacac 60
QY 1160 cactatgtccattttgaaacaaaagctcctcaacataccgcatccgacaaaattga 1219
DB 61 cactatgtccattttgaaacaaaagctcctcaacataccgcatccgacaaaattga 120
QY 1220 ggtcataaagccagagaagactatttcagttctgttaaaacttcctcctaaatgctatt 1279
DB 121 ggtcataaagccagagaagactatttcagttctgttaaaacttcctcctaaatgctatt 180
QY 1280 tgcgtcaagaagttccaca 1298
DB 181 tgcgtcaagaagttccaca 199

RESULT 9
AAC28094
ID AAC28094 standard; cDNA; 336 BP.
XX

Db 1080 -----|||||
 QY 1081 gtatcattacagagtagaagatgtcaacagctatctgtacatacttacagatcat 1140
 Db 1091 gtatcattacagagtagaagatgtcaacagctatctgtacatacttacagatcat 1150
 QY 1141 cagctattgagagacacacactatgtgccattttgaaacaaaagctcctcaacaatc 1200
 Db 1151 cagctattgagagacacacactatgtgccattttgaaacaaaagctcctcaacaatc 1210
 QY 1201 cgcctcgcagcaaaattgaggtcattataagccagagactatttcagctctgttaacct 1260
 Db 1211 cgcctcgcagcaaaattgaggtcattataagccagagactatttcagctctgttaacct 1270
 QY 1261 cattgcctaaatgctcattgtgtgcaagaagtccaca 1298
 Db 1271 cattgcctaaatgctcattgtgtgcaagaagtccaca 1308

RESULT 5

AAH05303
 ID AAH05303 standard; cDNA; 576 BP.

AC AAH05303;

XX 26-JUN-2001 (first entry)

XX Human cDNA clone (5'-primer) SEQ ID NO:2138.

DE Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 98JP-0248036.

XX 27-AUG-1999; 93JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -

XX Claim 1; SEQ ID 2138; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AA992446 to AA995893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

XX Sequence 576 BP; 167 A; 113 C; 114 G; 173 T; 9 other;

Query Match 42.0%; Score 545; DB 22; Length 576;

Best Local Similarity 98.7%; Pred. No. 1.5e-154;

Matches 545; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 atgtcttgggtccagcaacaattatataatataacacccctgaatcaactaaaggtggt 60

Db 25 atgtcttgggtccagcaacaattatataatataacacccctgaatcaactaaaggtggt 84

QY 61 acaattgtcaatgtctatgtgtgtggaagtcttttaagcccccatatcttaagcaaaagga 120

Db 85 acaattgtcaatgtctatgtgtgtggaagtcttttaagcccccatatcttaagcaaaagga 144

QY 121 actgattatgtcagttgttaactattgtgaccagacaaatgtataaactaacttgctg 180

Db 145 actgattatgtcagttgttaactattgtgaccagacaaatgtataaactaacttgctg 204

QY 181 ctcttagtgaaactatgaagcccttcccaataattataaaaatggagatattgttcgc 240

Db 205 ctcttagtgaaactatgaagcccttcccaataattataaaaatggagatattgttcgc 264

QY 241 ttccacaggtcgaagattcaagtataataaaaggagactcagggtatcaccagctctggc 300

Db 265 ttccacaggtcgaagattcaagtataataaaaggagactcagggtatcaccagctctggc 324

QY 301 ttgcatctttgacgtttgagggaactttggagccctcatcacctcgcacttcaagc 360

Db 325 ttgcatctttgacgtttgagggaactttggagccctcatcacctcgcacttcaagc 384

QY 361 aagtatttaacttcactactgagaccacacaaatggtagaagccttactgtttgggca 420

Db 385 aagtatttaacttcactactgagaccacacaaatggtagaagccttactgtttgggca 444

QY 421 tctactcatatgtcacgcttctggacattactataaaattgtgtgatgttcagccaatgcag 480

Db 445 tctactcatatgtcacgcttctggacattactataaaattgtgtgatgttcagccaatgcag 504

QY 481 tattttgacctgacttgcagctcttggcgaagcagaagtgacgagcatcattctt 540

Db 505 tattttgacctgacttgcagctcttggcgaagcagaagtgacgagcatcattctt 564

QY 541 cttaaaggtatgg 552

Db 565 cttaaaggtatgg 576

RESULT 6

AAH07649

ID AAH07649 standard; cDNA; 854 BP.

XX AC AAH07649;

XX 26-JUN-2001 (first entry)

XX Human cDNA clone (5'-primer) SEQ ID NO:4484.

XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -

Claim 8: SEQ ID 12760; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 2077 BP; 657 A; 401 C; 393 G; 626 T; 0 other;

Query Match 80.7%; Score 1048; DB 22; Length 2077;
Best Local Similarity 90.8%; Pred. No. 3.2e-306;
Matches 1178; Conservative 0; Mismatches 0; Indels 120; Gaps 1;

QY 1 atgtcttgggtccagcaaaaattatataatatacaccctgaatcaactaaagggtggt 60
DB 131 atgtcttgggtccagcaaaaattatataatatacaccctgaatcaactaaagggtggt 190
QY 61 acaattgtcaatgtcataggtgtgtgaagttctttaagccccccatataatgaagcaaga 120
DB 191 acaattgtcaatgtcataggtgtgtgaagttctttaagccccccatataatgaagcaaga 250
QY 121 actgattattgtcagttgaactatttggaccagacaaatgtaaaactaacttgcctg 180
DB 251 actgattattgtcagttgaactatttggaccagacaaatgtaaaactaacttgcctg 310
QY 181 ctctttagtggaactatgaagcccttccaaataattataaaaatggagatatgttcgc 240
DB 311 ctctttagtggaactatgaagcccttccaaataattataaaaatggagatatgttcgc 370
QY 241 ttccacaggtcgaagattcaagtataaaaaggagactcagggtatccaccgctctgac 300
DB 371 ttccacaggtcgaagattcaagtataaaaaggagactcagggtatccaccgctctgac 430
QY 301 ttgcatctttgacgtttgagggaactttggagcccttatcatcacctcgacattcaagc 360
DB 431 ttgcatctttgacgtttgagggaactttggagcccttatcatcacctcgacattcaagc 490
QY 361 aagatttttaacttcaactactagagaccacaaatgttagaagccttactgtattgggca 420
DB 491 aagatttttaacttcaactactagagaccacaaatgttagaagccttactgtattgggca 550
QY 421 tctactcatatgtcacctgtctggacattactaaaaattgtgtgattttcagccaatgacg 480
DB 551 tctactcatatgtcacctgtctggacattactaaaaattgtgtgattttcagccaatgacg 610
QY 481 tattttgacctgactgtcagctcttgggcaagcagaagtggaagcagcatcattttt 540
DB 611 tattttgacctgactgtcagctcttgggcaagcagaagtggaagcagcatcattttt 670
QY 541 ctaaaaggtatggatggcaccaggacacatttccattcttggagagtcttaatacaagac 600

DB 671 ctaaaaggtatggatggcaccaggacacacatttccattcttgagagtttatacaagac 730
QY 601 cttgtcttgaaggtgatttaagtcacatccatcggttcacaaaatctgacaatagacatt 660
DB 731 cttgtcttgaaggtgatttaagtcacatccatcggttcacaaaatctgacaatagacatt 790
QY 661 ttatgtctacgataaaccatgttcatgttgcaagaatctctgaaggttggagtttttttaga 720
DB 791 ttatgtctacgataaaccatgttcatgttgcaagaatctctgaaggttggagtttttttaga 850
QY 721 actatagccttcataccaaaacttcaatcaatgaattcagagaatcagacaatgttaagt 780
DB 851 actatagccttcataccaaaacttcaatcaatgaattcagagaatcagacaatgttaagt 910
QY 781 ttatagtttctatcttcaggtgaggtaccagttacggttcggtggaatcaggttcttcagaaa 840
DB 911 ttatagtttctatcttcaggtgaggtaccagttacggttcggtggaatcaggttcttcagaaa 970
QY 841 agtaactctgatgtggatacaactgaaaaggatttagaattcgaatttgacagccaat 900
DB 971 agtaactctgatgtggatacaactgaaaaggatttagaattcgaatttgacagccaat 1030
QY 901 cagcattcagatgttatctgtcaatcagaacctgacgacagcttcccaatggaggtctcg 960
DB 1031 cagcattcagatgttatctgtcaatcagaacctgacgacagcttcccaatggaggtctcg 1079
QY 961 ctctgtctccaggtcggagttcagtggaacggtctcgtcgtcatttgcagcctccacctcc 1020
DB 1080 ----- 1079
QY 1021 tgagttcaagcttctcgtcctcagcctcccaagtagctgggattacagggctctggatca 1080
DB 1080 -----gctctggatca 1090
QY 1081 gtatcattatacaggtgaaagatgtcaacagctatctgtcacaaatcttaccagatcat 1140
DB 1091 gtatcattatacaggtgaaagatgtcaacagctatctgtcacaaatcttaccagatcat 1150
QY 1141 cagatttggagagagacaccactatgtccattttgaacaaaagctcctcaacaatc 1200
DB 1151 cagatttggagagagacaccactatgtccattttgaacaaaagctcctcaacaatc 1210
QY 1201 cgcattccagcaaaattgaggttcataatgaagccagagactatttcagctgtttaaactt 1260
DB 1211 cgcattccagcaaaattgaggttcataatgaagccagagactatttcagctgtttaaactt 1270
QY 1261 cattgcccataatgtcatttgcgtcgaagaagttccaca 1298
DB 1271 cattgcccataatgtcatttgcgtcgaagaagttccaca 1308

RESULT 2

AAH17603

ID AAH17603 standard; cDNA: 2383 BP.

XX AAH17603;

AC AAH17603;

XX 26-JUN-2001 (first entry)

XX Human cDNA sequence SEQ ID NO:17112.

DE Human cDNA sequence SEQ ID NO:17112.

XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.

KW Homo sapiens.

OS Homo sapiens.

PN EP1074617-A2.

XX 07-FEB-2001.

PD 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 95JP-0248036.

PR

```

Oy 1126 atactacagatcatgatttgagagagacaccactatgtccatttttgaacaaaa 1185
Db 242 ATACTTACAGATCATGATTGAGAGGACACCATATGTGCCATTTTGAACAAAAA 301
Oy 1186 gctctcaacaataccgcacccagcaaaattgagggtcatataagcccaagaactattt 1245
Db 302 GCTCTCAACAATACCGCATCCGAGCAAAATTGAGGTCTATTAAGCCCAAGACATTT 361
Oy 1246 cagctgttaaaccttccatccctaaatgtca-ttctgctcaagaagttccaca 1298
Db 362 CAGTCTGTAAACTTTCATTCGCCCTAAATGTCATTTTCTGCTCAAGAAGTTCCACA 415

RESULT 15
AV656827
LOCUS AV656827 328 bp mRNA EST 07-SEP-2000
DEFINITION AV656827 GLC Homo sapiens cDNA clone GLCEVD06 3', mRNA sequence.
ACCESSION AV656827
VERSION AV656827.1 GI:9877841
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 328)
Xuan,B., Wu,I., Huang,Q., Kang,B., Gao,X., Xu,Z., Xiao,H.,
, Xu,X., Peng,Y., Liu,F., Ou,J., Song,H., Cheng,Z., Ou,J.,
Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
,G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.
Homo sapiens cDNA clone
Unpublished (2000)
Contact: Zequang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanqg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source
1..328
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GLCEVD06"
/clone_lib="GLC"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 103 a 58 c 55 g 111 t 1 others
ORIGIN

Query Match 20.6%; Score 268; DB 10; Length 328;
Best Local Similarity 100.0%; Pred. No. 9.4e-58;
Matches 268; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 atgtcttgggtccagcaacaattatatatacacccctgaatcaacttaagggtggt 60
Db 45 ATGCTTTGTTCCAGCAACAATATATATATACACCCCTGAATCAACTTAAAGGIGGT 104
Oy 61 acaattgtcaatgtctatggtgtgtggaagttctttaagcccccatatcttaagcaagga 120
Db 105 ACAATTGTCAATGCTATGTTGTTGTAAGTCTTTAAGCCCCCATATCTTAAGCAAGGA 164
Oy 121 actgattattgctcagttgtaactatttgaggaccagacaaatgtaaaactaactgcctg 180
Db 165 ACTGATTATTGCTCAGTTGTAATCTATTGTGGACACAGACAAATGTAAACTAACITGCCTG 224
Oy 181 ctctttagtggaactatgaagccctcccaataattataaaaatggagatatttctgcg 240
Db 225 CTCTTTAGTGAACACTATGAAGCCCTTCCCAATANTTTATAAAAATGGAGATATTGTCOC 284

```

```

Oy 241 ttccacaggtcgaagattccaagtatafa 268
Db 285 TTTCACAGGCTGAAGATTCAAGTATATA 312

```

Search completed: April 9, 2002, 22:13:58
Job time: 8873 sec

pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcggg
); Oligo-dr primed using primer 5'-TTTTTTTTTTTTTTN-3',
size-selected for average insert size 2.2 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NHGRI, National
Institutes of Health). Note: this is a NIH-MGC Library.*

BASE COUNT 302 a 197 c 205 g 299 t 1 others
ORIGIN

Query Match 24.4%; Score 317.2; DB 11; Length 1004;
Best Local Similarity 83.6%; Pred. No. 3.5e-70;
Matches 443; Conservative 0; Mismatches 78; Indels 9; Gaps 7;

Qy 1 atgtcttgggtccagcaacaattatatatacacccctggaactcaacttaagggtggt 60
|||||
Db 432 ATGCTTTGGTCCAGCAACAATATATATATACACCCCTGAATCACTTAAGGGIGGT 491
|||||
Qy 61 acaattgtcaatgtctatggtgtgtgaagttctttaagcccccatatataagcaaaagga 120
|||||
Db 492 ACAATGTCAATGTCTATGGTGTGTGAAGTCTTTTAAGCCCCCATATCTAAGCAAGGA 551
|||||
Qy 121 actgattatgtcagttgaactattgtggaccagacaaaatgtaaaactaacttgcctg 180
|||||
Db 552 ACTGATTATGCTCAGTGTGAATGTGGACGACGACAAATGT-AACTAACITGCCTG 610
|||||
Qy 181 ctcttttagtgaactatgaagccctcccaataattata--aaatggagatattgttc 238
|||||
Db 611 CTCTTTAGTGAACATATGAAGCCCTTCCTCAATTAATTAACACAAATGGAGATATGTTT 670
|||||
Qy 239 gcttcacaggtggaattcaagtataataaaggagactcaggtatcacagctctg 298
|||||
Db 671 GCTTTACAGGCTGAGATTCAGTATATTAACGAGACTCA-GGTATCACCAGTCTG 729
|||||
Qy 299 gctttgcatctttgacgtttgagggaactttggagccctatcatcacctcgacctcaa 358
|||||
Db 730 GCTTTGCATC-TTGAGCTGTAAGAACTTGGGAGCCCTATCATAACTCGGCATTTCA 788
|||||
Qy 359 gcaagtatttaacttactactagga--ccacaaatggttagaa--gccttactgttt 415
|||||
Db 789 GCCAGTATTAACTTCCTTACTGTAGGACCCACAAATGGTAGAAGGCCCTTCGTCGTG 848
|||||
Qy 416 gggcatctactatgtcaccgctcttgacattactaaattgtgtgattgttcagccaa 475
|||||
Db 849 GGGCATCTACCCATATGATACCGGCTCGCAATTTCTCACA-TGTTGGAGTTCAGCACA 907
|||||
Qy 476 tgcagtatttgacctgaacttgcagctcttgggcaaaagcagaagtggac 525
|||||
Db 908 GGACGTATAGAAACGAGATTGTGCCCCGTGCAAGAAATGTCAGAGGAC 957
|||||

RESULT 12
LOCUS AA382297 341 bp mRNA EST 21-APR-1997
DEFINITION EST95690 Testis 1 Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION AA382297
VERSION AA382297.1 GI:2034796
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ADAMS, M.D., SOARES, M.B., KERLAVAGE, A.R., FIELDS, C. and VENTER, J.C.
Rapid cDNA sequencing (expressed sequence tags) from a
directionally cloned human infant brain cDNA library;
NATURE GENET. 4, 373-380 (1993)
JOURNAL 94004965
MEDLINE
COMMENT Other ESTs: THCL21679
Contact: Kerlavage, AR
Bioinformatics

FEATURES
source
1..341

/organism="Homo sapiens"
/db_xref="ATCC (inhost):186770"
/db_xref="taxon:9606"
/clone_lib="Testis 1"
/sex="male"
/dev_stage="adult"
/note="Organ: testis; Vector: pBluescript SK-; Site_1:
ECORI; Site_2: XhoI"

BASE COUNT 108 a 68 c 67 g 98 t
ORIGIN

Query Match 24.0%; Score 311.8; DB 10; Length 341;
Best Local Similarity 99.1%; Pred. No. 6.8e-69;
Matches 324; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 625 cacatccatcaggtacaaaatctgacatagacatttttagtctacgataaccatgttc 684
|||||
Db 1 CACATCCATCGGCTACAAATCTGACATAGACATTTTAGTCTAGATAACCATGTTGAT 60
|||||
Qy 685 gtggcaagatctctgaagttggaagctttcttagaattctatagccttcataccaaactt 744
|||||
Db 61 GTGGCAAGATCTCTGAAGTTTGGAAAGCTTCTTAGAATCTATAGCCTTCATACCAACTT 120
|||||
Qy 745 caatcaatgaattcagagaatcagacaattgtaatttagagtttcattcttcagaggt 804
|||||
Db 121 CAATCAATGAATTCAGAGAATCAGACAATGTTAAGTTTAGAGTTTCACTTCATGGAGGT 180
|||||
Qy 805 accagttacggtcgggggaatcagggtcttggcagaaagtaactctgattgtgatacaactg 864
|||||
Db 181 ACCAGTTACGGTCGGGGAATCAGGCTCTGCCAGAAAGTAACCTGATGTGATCAACTG 240
|||||
Qy 865 aaaaaggatttagaactctgaaaatttgacagccaa-tcagcattcagattgtattgtca 923
|||||
Db 241 AAAAAGGATTAGAACTCTGCAAAATTGACGCAATTCAGCAATTCAGATGTTATCTGTCA 300
|||||
Qy 924 atcagaacctgacgacagctttccaaa 950
|||||
Db 301 ATCAAACTGACGACAGCTTTTCCAA 327
|||||

RESULT 13
LOCUS BG088861 565 bp mRNA EST 26-JAN-2001
DEFINITION H3158E09-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
ACCESSION BG088861
VERSION BG088861.1 GI:12571425
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ADAMS, M.D., SOARES, M.B., KERLAVAGE, A.R., FIELDS, C. and VENTER, J.C.
Rapid cDNA sequencing (expressed sequence tags) from a
directionally cloned human infant brain cDNA library;
NATURE GENET. 4, 373-380 (1993)
JOURNAL 94004965
MEDLINE
COMMENT Other ESTs: H3158E09-3
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health

The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.
Location/Qualifiers


```

Db 418 ATCAATGACATCCAGAGATCAGACAA-GTTAACGTCAGAGTCTCATCTTCATGGAGG 476
Oy 804 taccagtacgggtcggaatcacagggtcttgcagaaa 841
Db 477 TACCAGTTACGGTCGGGGCATCAGGGTCTTGCCAAAGAA 514

RESULT 8
AUI24189 866 bp mRNA EST 23-OCT-2000
LOCUS AUI24189 NT2RM2 Homo sapiens cDNA clone NT2RM2001805 5', mRNA
DEFINITION sequence.
ACCESSION AUI24189
VERSION AUI24189.1 GI:10948905
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 866)
AUTHORS Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S., Isogai,T.)
TITLE HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki
,Y., Sugano,S., Isogai,T.)
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source
1. .866
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RM2001805"
/clone_lib="NT2RM2"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
precursor cells"
BASE COUNT 254 a 162 c 171 g 274 t 5 others
ORIGIN
1 atgtcttgggtccagcaacaattatatatatacaccctgaatcaacttaagggtggt 60
439 ATGCTTTGGTCCACACAAATATATATATACACCCCTGAATCAACTTAAGGGTGGT 498
Oy 61 acaattgtcaatgtctatggtgttggaagttctttaagcccccatatcttaagcaagga 120
Db 499 ACAATTGTCAATGCTCTATGGTGTGTGAAGTCTTTAAGCCCCCATATCTTAAGCAAGGA 558
Oy 121 actgattattgtcagttgaactatttggaccgagacaataatgtaaaactaactgcctg 180
Db 559 ACTGATTATGCTCAGTTGTAAGTATTTGGACCAACAAATGTAAAACTTAACTTTCCTG 618
Oy 181 ctctttagtggaactatgaagcccttccaataattataaaatggagatatgttcgc 240
Db 619 CTCTTTAGTGGAACTATGAAGCCCTTCCAATAATTATATAAAATGGAGATATTGGTCGC 678

```

```

Oy 241 ttccacagggtgaagattcaagtataataaaagagagactcagggtatcaccagctctggc 300
Db 679 TTTACAGGCTCAAGATTCAAGTNTAAAAAGAGAGACTCAGGGTATCACCAGCTCTGGC 738
Oy 301 ttgcatctttgaagtttgagggaactttggagagccctatcataccctcgcaactcaagc 360
Db 739 TTTGCACTCTTTGACGTT--GAGGGACTTTGGGAGCCCCCTATCATACCTTGCACTTNAGCC 796
Oy 361 aagatttttaacttccactactgagagaccacaaaatgtagaagccttcagtggttgggca 420
Db 797 AGTATTTTACTTCTTACT-----GAGGCCNCAATGGNAGAAGCCTTACGGGTTTGGGCT 851
Oy 421 t 421
Db 852 t 852

RESULT 9
AUI24189 409 bp mRNA EST 15-OCT-1999
LOCUS AUI24189 NT2RM2 Homo sapiens cDNA clone
DEFINITION IMAGE:2581377 3', mRNA sequence.
ACCESSION AUI24189
VERSION AUI24189.1 GI:6043420
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 409)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 403.
FEATURES
source
1. 409
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2581377"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI-CGAP_GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 120 a 79 c 74 g 136 t
ORIGIN
1
Query Match 25.9%; Score 336.6; DB 10; Length 409;
Best Local Similarity 98.8%; Pred. No. 3.5e-75;
Matches 339; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Oy 609 tgaagggtattgaagtcacatccatcgctcacaaaatctcacaaatagacatttagtcta 668
Db 409 TGAAGGTGATTTAAGTACATCCCATCGCTACAAAATCTCACAATAGACATTTAGTCTA 350
Oy 569 cgataaccatgttcatgtggcaagatctctgaaggttggagcttcttctagaatctatag 728

```

/clone="PLACE1005026"
/clone_lib="PLACE1"
/tissue_type="placenta"
/note="Vector: PME18SFL3"
259 a 164 c 166 g 259 t 6 others

BASE COUNT
ORIGIN

Query Match 40.2%; Score 522.4; DB 10; Length 854;
Best Local Similarity 99.3%; Pred. No. 2.2e-122;
Matches 545; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1 atgtcttgggtccagcaacaataatatatatatacacccctgaatcaacttaaggttgg 60
Db 132 ATGTCTTTGGTTCAGCAACAATAATATATATATACACCCCTGAATCAACTTAAGGTTGT 191
QY 61 acaattgcaatgtctatggtgtgtgaagttctttaagcccccatacttaagcaagga 120
Db 192 ACAATTTGCAATGTCTATGTTGTGTGAAGTTCTTTAAGCCCCCATAATCTTAAGCAAGGA 251
QY 121 actgattattgtctagttgtaactattgtggaccagacaaatgtataaactaacttgcctg 180
Db 252 ACTGATTATTGCTCAGTTGTAACATTGTGGACGACACAAATGTAAACTTAACCTTGCCCTG 311
QY 181 ctctttagtgaaactatgaagcccttccaataattataaaatggagatatattgtcgc 240
Db 312 CTCTTTAGTGGAACTATGAAGCCCTTCCAATAATTTATAAANAATGGAGATATTGTTCGC 371
QY 241 ttccagagctgaagattcaagtataataaaagagactcagggtatcacccagctctggc 300
Db 372 TTTCACAGCTGAGATTCAGTATATAAAGAGGACATCAGGGTATACACGACTTCGGC 431
QY 301 ttgcatctttgacgtttgagggaaactttgggagcccttatcatcactgcacttcaagc 360
Db 432 TTTCATCTTTGACGTTTGAGGAACTTTGGAGGCCCTTATCATACCTCGCATTCACGC 491
QY 361 aagatttttaacttcactactagacacacaaaatggtagaagccttactgtttgggca 420
Db 492 AGTATTTTACTTCTACTACTGAGACCCACAAAATGGTAGAGCCTTACTGTTTGGGCA 551
QY 421 ttactcatatgtccacctgttggacattactataaattgtgtgtgttcagcc-aatgca 479
Db 552 TCTACTCATATGTCACCGTCTTGGACATTACTATAAATTTGTGTGTGTTTTCAGCCAAATGCA 611
QY 480 gtattttgacctgactgtgactcttggcgaagcagaagtgagc-gagcatcatttc 538
Db 612 GTATTITGACCTGACTTGCAACTCTTGGCAAGCAAGGTGGAGGNGCATCAITTC 671
QY 539 ttctaaagg 547
Db 672 TTCTAAAGG 680

RESULT 5
AA351459/c 429 bp mRNA EST 21-APR-1997
LOCUS EST59213 Infant brain Homo sapiens cDNA 5' end, mRNA sequence.
DEFINITION AA351459
ACCESSION AA351459
VERSION AA351459.1 GI:2003779
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 429)
AUTHORS Adams,M.D., Soares,M.B., Kerlavage,A.R., Fields,C. and Venter,J.C.
TITLE Rapid cDNA sequencing (expressed sequence tags) from a
directionally cloned human infant brain cDNA library
JOURNAL Nature Genet. 4, 373-380 (1993)
MEDLINE 94004965
COMMENT Other ESTs: EST59212 THC121679
Contact: Kerlavage, AR
Bioinformatics

The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
for clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.

FEATURES
Source

1. 429
/organism="Homo sapiens"
/db_xref="ATCC (Inhost):151767"
/db_xref="taxon:9606"
/clone_lib="Infant brain"
/sex="female"
/dev_stage="Infant"
/note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
Site_2: NotI"
125 a 94 c 84 g 125 t 1 others

BASE COUNT 125 a 94 c 84 g 125 t 1 others
ORIGIN

Query Match 29.7%; Score 385.2; DB 10; Length 429;
Best Local Similarity 97.2%; Pred. No. 1.5e-87;
Matches 413; Conservative 0; Mismatches 9; Indels 3; Gaps 2;

QY 399 aagaagccttactgtttgggcatc--tactcataatgtcacctgttggac-attactaaa 455
Db 425 ACAAGCCTTACGTTGGGGCAACCTACTCCATATGNCACCGCTCTGGGACAATTACTAAA 366
QY 456 attgtgtgattgttcagccaatcagattttgaactgacttgcagctcttgggcaaacg 515
Db 365 ATTGTGTATGTTCAGCCCAATGCAGTATTTTACCTGACTTGTGACGCTCTGGGAAAGC 306
QY 516 aagaagtgacggagcatcatttctctaaagtgatggatggcaccagacacatttcc 575
Db 305 AGAAGTGGACGGAGCATCATTTCTTAAAGGTATGGATGGCACCAGGACCATTTAC 246
QY 576 atcttggagagctcttaatacaagacctgttcttgaagggtatttaagtcacatccatcg 535
Db 245 ATCTTGGAGAGCTTTAATACAAGACCTTGTCTTGAAGGTGATTTAAGTCACATCCATCG 186
QY 636 gctcaaaatctgacaatagacatttttagtctacgataaccatgttcatgtggcaagatc 695
Db 185 GCTACAAATCTGACAATAGACATTTTAGCTACGATAACCATGTTCTATGTGCAAGATC 126
QY 696 tctgaaggttggagctttcttagaatctatagccttcataccaaacttcaatcaatgaa 755
Db 125 TCTGAAGGTTGGAGCTTTCTTTAGAATCTATAGCCTTCATACCAACTTCAATCAATGAA 66
QY 756 ttcagagaatcagacaatgttaagtttagagtttcatcttcatgaggtaccagttacgg 815
Db 65 TTCAGAGAAATCAGACAAATGTTAAGTTTAGAGTTTCTATCTTTCATGAGGTACCATGACG 6

RESULT 6

AL520360 782 bp mRNA EST 13-FEB-2001
LOCUS AL520360 LTI_NFL004_NBC2 Homo sapiens cDNA clone CS0DB006YH17 5
DEFINITION prime, mRNA sequence.
ACCESSION AL520360
VERSION AL520360.1 GI:12783853
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 782)

/note="Organ: skin; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT 271 a 205 c 212 g 238 t
ORIGIN

Query Match 47.1%; Score 611.2; DB 11; Length 925;
Best Local Similarity 94.0%; Pred. No. 5.8e-145;
Matches 679; Conservative 0; Mismatches 36; Indels 5; Gaps 4;
QY 87 gaagttcttaagcccccatactcaagaaagaaactaattattgctcagttgtaactat 146
|||||
Db 1 GAAGTTCTTTAGCCCCCAATCTATAGCAAGAACTGATTATTGCTCAGTTGTAACTAT 60
QY 147 tctggaccagacaaatgtaaaactaaactgctgctctcttadtgtgaaactatgaagccct 206
|||||
Db 61 TGTGGACCAGACAAATGTAAACTAACTTTCCTGCTCTTTAGTGGAACTATGAGGCCCT 120
QY 207 tcaataattataaaatggagatattgttcgcttccacaggtgaagattcaagtata 266
Db 121 TCCAAATAATTTTAAATAATGGAGATATTGTTCGCTTTACAGGCTGAAGATCAAGTATA 180
QY 267 taaaaggagactcagggtatcaccagctctgcttgcattcttgacgtttgaagaaac 326
|||||
Db 181 TAAAAGGAGACTCAGGGTATCACAGCTCTGCTTTGATCTTTGACGCTTGGAGGAAAC 240
QY 327 tttggagccctatacatacctcgcaactcaagcaagatattttaacttcaactgaagga 386
Db 241 TTTGGAGCCCTATCATACCTCGCACTTCAAGCAAGTATTTAACTTCACTACTAGCA 300
QY 387 ccacaaatggtagaacctgagttgtttgggcatctactataatgtaacgtcttgac 446
Db 301 CCACAAATGGTAGAGCCCTTACGTTTGGGCATCTACTATATGTACCGCTCTGGAC 360
QY 447 attactaaaattgtgtatgtttcagccaatgcagtatatttggacctgaactgtcagctctt 506
Db 361 ATTACTAAAATTTGTGTATGTTTGCCTTTCAGCAATGCAGTATTTTGACCTGACTTGTACGCTCT 420
QY 507 gggcaagcagaagtggagcagagatcattttcttaaggtatggatggcaccagac 566
Db 421 GGGCAAGCAGAAGTGGAGCGAGCATCATTTCTTAAAGGTATGGATGGCACCAGGAC 480
QY 567 accatttccatcttgagagctcttaacaagacctgttcttgaagtgatttaagtca 626
Db 481 ACCATTTCATCTGGAGAGCTTAATACAGACCTTGTCTTGAAGGTGATTTAAGTCA 540
QY 627 catcatcggtcacaaatctgacaatagacatttttagtctacgataaaccatgttcatgt 686
Db 541 CATCCATCGG-TACAAAATCTGACAAATAGACATTTTAGTCTACGATAACCATGTTCATGT 599
QY 687 g-gcaagatctcgaaggttgaagcttcttagaactctatagctctcataccaaacttc 745
Db 600 GCGCCAGATCTCTGAAGGTGGAAG-ATCTTAGAATCTATATGCTCTTAGCCAA--CTT 556
QY 746 atcaatgaattcagagaatcagacaatgttaagtttagagtttcatcttcatgtgaggtta 805
Db 657 CATCATGGAATTCGGAGAAATCGGACACGGTTTCAGTTTTCAGGTCATCTCCAGGAGGCCA 716
QY 806 CC 807
Db 717 CC 718

RESULT 2
BF027455
LOCUS BF027455 920 bp mRNA EST 10-OCT-2000
DEFINITION 601672751F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3955573 5',
mRNA sequence.
ACCESSION BF027455
VERSION BF027455.1 GI:10735167
KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 920)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rcgaps-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DP

CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: LLCMB30 row: e column: 14

High quality sequence start: 52

High quality sequence stop: 746.

Location/Qualifiers

1. 920

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3955573"

/clone_lib="NIH_MGC_20"

/tissue_type="melanotic melanoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:

ECORI; CDNA made by oligo-dr priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGCACGAG(G). Size-selected >500bp for average

insert size 1.8kb. Library constructed by Ling Hong in

the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 275 a 189 c 189 g 267 t

ORIGIN

Query Match 42.3%; Score 548.6; DB 11; Length 920;
Best Local Similarity 92.0%; Pred. No. 4.9e-129;
Matches 702; Conservative 0; Mismatches 44; Indels 17; Gaps 11;
QY 7 ttgttccagcaacaattatatatacacccctgaatcaacttaaggtgggtacaatt 66
Db 97 TTGGTCCAGCAACAATATATATATACACCCCTGAATCAACTTAAGGTGGTACAA-T 155
QY 67 gtcaatgtctatggtgtgtgaagttctttaagcccccatatctaaagcaagaaactgat 126
Db 156 GTCAATGTCTATGTTGTGAAGTTCTTTAAGCCCCCATATCTTAAGCAAGGAACGTAT 215
QY 127 tattgtcagttgaactattgtgaccagacaataataaaactaaacttgcctgctctt 186
Db 216 TATTGCTCAGTTGTAACTATTGTGGACAGACAAATGTAAACTTAACCTTGCTGCTCTTT 275
QY 187 agtggaaactatgaagcccttccaataattataaaatggagatatgttgcgtttcac 246
Db 276 AGTGAAACTATGAAGCCCTTCCAATAATTATATAAATGGAGATATTGTTCGCTTTAC 335
QY 247 aggtgaagattcaagtataataaagagagactcaggggtatcaccagctctggtttgca 306
Db 336 AGGCTGAAGATTCAAGTATATAAAGAGAGACTCAGGGTATCACACGCTCTGGCTTTGCA 395
QY 307 tctttgacgtttgaggaactttggagccctcatatacctcgcacttcaagcaag-ta 365
Db 396 TCTTTGAGTT--GAGGGAACCTTGGAGCCCTATCATACCTCGCACTTCAGCAAGCTA 453
QY 366 ttttaacttactactaggagaccacaaatgttagagccttacctgtgtttgggcatctac 425
Db 454 TTTTAACTTCACTACTAGGACCAACAAATGGTAGAAGCCCTTAGCTG-TTGGGCACTAC 512
QY 426 tcatatgtcaccgtctggacattactaaattgtgtgatttcagccaatgcagattt 485
|||||

```
Query Match          6.7%; Score 87; DB 4; Length 10380;
Best Local Similarity 83.5%; Pred. No. 1.8e-17;
Matches 111; Conservative 0; Mismatches 20; Indels 2; Gaps 1;

QY 943 ttccaatgagctccttcgctcagcagctgagttcagtcagtcgagtcgctc 1002
DB 6794 TTTGAGATGAGATCTCGCTCTGTCCACCCAGGCTGGAGTGCAGTCGATCTCAGTTC 6853
QY 1003 attgcagctccaccctcctgagttcaagc--ttctcctgcctcagcctcccaagttagctg 1060
DB 6854 ACTGCAACCTTGTCTCTGAGTTCAAGCGATTCTCTCTGCTCAGCCTCCTGAGTAGCTG 6913
QY 1061 ggattacagctc 1073
DB 6914 GGATTACAGCGC 6926

RESULT 13
US-09-173-914-6
; Sequence 6, Application US/09173914
; Patent No. 6171857
; GENERAL INFORMATION:
; APPLICANT: Hendrickson, Eric
; TITLE OF INVENTION: A No. 6171857el Leucine Zipper, KARP-1 and
; TITLE OF INVENTION: Methods of Regulating DNA Dependent Protein Kinase Activity
; FILE REFERENCE: B0877/7017/HK
; CURRENT APPLICATION NUMBER: US/09/173,914
; CURRENT FILING DATE: 1998-10-16
; EARLIER APPLICATION NUMBER: 60/064,557
; EARLIER FILING DATE: 1997-10-17
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 14636
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-173-914-6

Query Match          6.7%; Score 87; DB 4; Length 14636;
Best Local Similarity 83.5%; Pred. No. 2.2e-17;
Matches 111; Conservative 0; Mismatches 20; Indels 2; Gaps 1;

QY 943 ttccaatgagctccttcgctcagcagctgagttcagtcagtcgagtcgctc 1002
DB 8033 tttagacagagctcgcctctgtctgctccagcagctgagtgaaagtcgcgctcgcctc 8092
QY 1003 attgcagctccaccctcctcagttcaagc--ttctcctgcctcagcctcccaagttagctg 1060
DB 8093 actgcaacctccgcctcccccgggttcaagctattctcctcagcctcccccagtagctg 8152
QY 1061 ggattacagctc 1073
DB 8153 ggattacagcgc 8165

RESULT 14
US-09-345-217-3/c
; Sequence 3, Application US/09345217
; Patent No. 6268142
; GENERAL INFORMATION:
; APPLICANT: DUFF, GORDON W.
; APPLICANT: COX, ANGELA
; APPLICANT: CAMP, NICOLA J.
; APPLICANT: DIGIOVINE, FRANCESCO S.
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS FOR DISEASES ASSOCIATED
; TITLE OF INVENTION: WITH AN IL-1 INFLAMMATORY HAPLOTYPE
; FILE REFERENCE: MSA-010.02
; CURRENT APPLICATION NUMBER: US/09/345,217
; CURRENT FILING DATE: 1999-06-30
; EARLIER APPLICATION NUMBER: PCT/GB98/01481
; EARLIER FILING DATE: 1998-05-21
; EARLIER APPLICATION NUMBER: 9711040.7
```

```
; EARLIER FILING DATE: 1997-05-29
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 12565
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-345-217-3

Query Match          6.7%; Score 86.8; DB 4; Length 12565;
Best Local Similarity 85.2%; Pred. No. 2.3e-17;
Matches 109; Conservative 0; Mismatches 17; Indels 2; Gaps 1;

QY 946 ccaaatgagctccttcgctcctccagcagctgagttcagtcagtcgagtcgctcatt 1005
DB 10822 CGAGACGGAGTCTTGTCTGTGCGCTAGGCTGGAGTGCAGTGCACAAATCTCGGCTCACT 10763
QY 1006 gcagcctccaccctcctgagttcaagc--ttctcctgcctcagcctcccaagttagctggga 1063
DB 10762 GCAACCTCCACCTCTGGGTTCAAGCGATTCTCTCTGCTCAGCCTCCTGAGTAGCTGGA 10703
QY 1064 ttacaggc 1071
DB 10702 TTACAGGC 10695

RESULT 15
US-09-385-982-358
; Sequence 358, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 358
; LENGTH: 619
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(619)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-358

Query Match          6.7%; Score 86.6; DB 4; Length 619;
Best Local Similarity 84.0%; Pred. No. 4.3e-18;
Matches 110; Conservative 0; Mismatches 19; Indels 2; Gaps 1;

QY 943 ttccaatgagctccttcgctcctccagcagctgagttcagtcagtcgagtcgctc 1002
DB 29 tttagagatgagctcgcctctgtctgctccagcagctgagtcgagtcgagtcgcaatctcgtc 88
QY 1003 attgcagctccaccctcctcagttcaagc--ttctcctgcctcagcctcccaagttagctg 1060
DB 89 actgcaacctccgcctccctccgggttcaagcaattctcctcagcctccccaagttagctg 148
QY 1061 ggattacaggc 1071
DB 149 ggattacagcgc 159
```

```
FEATURE:
NAME/KEY: intron
LOCATION: 15116..15533
OTHER INFORMATION: /number= 11
FEATURE:
NAME/KEY: exon
LOCATION: 15534..15757
OTHER INFORMATION: /number= 12
FEATURE:
NAME/KEY: intron
LOCATION: 15758..16949
OTHER INFORMATION: /number= 12
FEATURE:
NAME/KEY: exon
LOCATION: 16950..17082
OTHER INFORMATION: /number= 13
FEATURE:
NAME/KEY: intron
LOCATION: 17083..17696
OTHER INFORMATION: /number= 13
FEATURE:
NAME/KEY: exon
LOCATION: 17697..17764
OTHER INFORMATION: /number= 14
FEATURE:
NAME/KEY: intron
LOCATION: 17765..18534
OTHER INFORMATION: /number= 14
FEATURE:
NAME/KEY: CDS
LOCATION: join(1776..1854, 2564..2621, 4076..4208,
6041..6252, 6802..6934, 7759..7856, 9444..9573,
10867..11081, 12481..12613, 13702..13799,
14977..15115, 15534..15757, 16950..17082,
17697..17741)
OTHER INFORMATION: /product= "human serum albumin"
OTHER INFORMATION: /citation= {[1]}
FEATURE:
NAME/KEY: exon
LOCATION: 18535..18697
OTHER INFORMATION: /number= 15
FEATURE:
NAME/KEY: 3'UTR
LOCATION: 17742..18697
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1737..1775
PUBLICATION INFORMATION:
AUTHORS: Minghetti, P P
AUTHORS: Ruffner, D E
AUTHORS: Kuang, W-J
AUTHORS: Dennison, O E
AUTHORS: Hawkins, J W
AUTHORS: Beattie, W G
AUTHORS: Dugaiczky, A
TITLE: MOLECULAR STRUCTURE OF THE HUMAN ALBUMIN
TITLE: GENE IS REVEALED BY NUCLEOTIDE SEQUENCE WITHIN
TITLE: q11-22 OF CHROMOSOME 4
JOURNAL: J. Biol. Chem.
VOLUME: 261
PAGES: 6747-6757
DATE: 1986
RELEVANT RESIDUES IN SEQ ID NO: 36: FROM 1 TO 19011
US-08-310-356-36
```

```
Query Match          6.7%; Score 87.2; DB 1; Length 19011;
Best Local Similarity 84.6%; Pred. No. 2.2e-17;
Matches 110; Conservative 0; Mismatches 18; Indels 2; Gaps 1;

Oy 943 ttccaagtgcgtcttcgtccctccaggctgagttcagtcgacgggtctcgctc 1002
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15309 TTTTGACGCGAGTCTCGCTTTGTTGTCAGGCTGGAGTGGTGCATCTCGCTC 15368
```

```
Oy 1003 attgcagctccacctcctgagttcaagc--ttctctgcctcagcctcccaagtagctg 1060
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15369 ACTGCAACCTCCGCTCCCAAGTTCAAGCCATTCTCTGCTCAGCTCCCAAGTAGCTG 15428

Oy 1061 ggattacagg 1070
||| | | | | |
Db 15429 GGACTACAGG 15438

RESULT 10
PCT-US92-06300-1
: Sequence 1, Application PC/TUS9206300
: GENERAL INFORMATION:
: APPLICANT: Hurwitz, David R
: APPLICANT: Nathan, Margret
: APPLICANT: Shani, Moshe
: TITLE OF INVENTION: Transgenic Protein Production
: NUMBER OF SEQUENCES: 1
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Rhone-Poulenc Rorer, Inc.
: STREET: 500 Virginia Ave., Bldg. 3A
: CITY: Ft. Washington
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19034
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US92/06300
: FILING DATE: 19920730
: CLASSIFICATION: 800
: ATTORNEY/AGENT INFORMATION:
: NAME: Goodman, Rosanne
: REGISTRATION NUMBER: 52,534
: REFERENCE/DOCKET NUMBER: A0856-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 962-4130
: TELEFAX: (215) 962-4107
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 19557 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: PUBLICATION INFORMATION:
: AUTHORS: Minghetti, P P
: AUTHORS: Ruffner, D E
: AUTHORS: Kuang, W.-J.
: AUTHORS: Dennison, O E
: AUTHORS: Hawkins, J W
: AUTHORS: Beattie, W G
: AUTHORS: Dugaiczky, A
: TITLE: MOLECULAR STRUCTURE OF THE HUMAN ALBUMIN GENE
: TITLE: IS REVEALED BY NUCLEOTIDE SEQUENCE WITHIN Q11-22
: TITLE: OF CHROMOSOME 4
: JOURNAL: J. Biol. Chem.
: VOLUME: 261
: PAGES: 6747-6757
: DATE: 1986
: RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 19002
PCT-US92-06300-1
```

```
Query Match          6.7%; Score 87.2; DB 5; Length 19557;
Best Local Similarity 84.6%; Pred. No. 2.2e-17;
Matches 110; Conservative 0; Mismatches 18; Indels 2; Gaps 1;
```


QY 1008 agctccacctctcgtagttcaagc--ttctctcgtcctcagcgtcccaagtagctaggatt 1065
|:|||||
Db 221 ARCTCCACCTCTGGTTCAAGCGATTCTCTGCTCAGGCTCCCGAGTAGCTGGGATT 162

QY 1066 acaggctc 1073
|:|||||
Db 161 ACAGGCGC 154

RESULT 2

US-08-724-394A-20/C
; Sequence 20, Application US/08724394A
; Patent No. 5872237

GENERAL INFORMATION:

APPLICANT: Feder, John N.

APPLICANT: Kronmal, Gregory S.

APPLICANT: Lauer, Peter M.

APPLICANT: Ruddy, David A.

APPLICANT: Thomas, Winston

APPLICANT: Tsuchihashi, Zenta

APPLICANT: Wolff, Roger K.

TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el

TITLE OF INVENTION: Sequences and Antibodies Thereto

NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESS:

ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP

STREET: Two Embarcadero Center, 8th Floor

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94111-3834

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/724,394A

FILING DATE: 01-OCT-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Fitts, Renee A.

REGISTRATION NUMBER: 35,136

REFERENCE/DOCKET NUMBER: 017957-000100

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-576-0200

TELEFAX: 415-576-0300

INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:

LENGTH: 246240 base pairs

TYPE: nucleic acid

STRANDEDNESS: not relevant

TOPOLOGY: not relevant

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1..246240

OTHER INFORMATION: /note= "HLA-H.CONTIG"

US-08-724-394A-20

Query Match 6.9%; Score 89.6; DB 2; Length 246240;
Best Local Similarity 84.3%; Pred. No. 1.7e-17;
Matches 113; Conservative 0; Mismatches 19; Indels 2; Gaps 1;

QY 943 ttcccaaatggagctcgtctcgtctccaggtcaggtcaggtcaggtcaggtcaggtc 1002
|:|||||
Db 8727 TTTTGACGGAGTCTCGTTCTGTCTACCCAGGCTGGAGTGCAGTGGCACAATCTCGGCTC 8668

QY 1003 attcagctccaccctcctcagttcaagc--ttctcctcctcagcgtcccaagtagctg 1060
|:|||||
Db 8667 ACTGCAACCTCCCGGTTCAAGCTATTCTCTCGCTCAGGCTCCCAAGTAGCTG 8608

QY 1061 ggattacaggctct 1074
|:|||||
Db 8607 GGACTACAGGCGCT 8594

RESULT 3

US-08-724-394A-21/C
; Sequence 21, Application US/08724394A
; Patent No. 5872237

GENERAL INFORMATION:

APPLICANT: Feder, John N.

APPLICANT: Kronmal, Gregory S.

APPLICANT: Lauer, Peter M.

APPLICANT: Ruddy, David A.

APPLICANT: Thomas, Winston

APPLICANT: Tsuchihashi, Zenta

APPLICANT: Wolff, Roger K.

TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el

TITLE OF INVENTION: Sequences and Antibodies Thereto

NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESS:

ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP

STREET: Two Embarcadero Center, 8th Floor

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94111-3834

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/724,394A

FILING DATE: 01-OCT-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Fitts, Renee A.

REGISTRATION NUMBER: 35,136

REFERENCE/DOCKET NUMBER: 017957-000100

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-576-0200

TELEFAX: 415-576-0300

INFORMATION FOR SEQ ID NO: 21:

SEQUENCE CHARACTERISTICS:

LENGTH: 246240 base pairs

TYPE: nucleic acid

STRANDEDNESS: not relevant

TOPOLOGY: not relevant

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1..246240

OTHER INFORMATION: /note= "HLA-H.CONTIG"

US-08-724-394A-21

Query Match 6.9%; Score 89.6; DB 2; Length 246240;
Best Local Similarity 84.3%; Pred. No. 1.7e-17;
Matches 113; Conservative 0; Mismatches 19; Indels 2; Gaps 1;

QY 943 ttcccaaatggagctcgtctcgtctccaggtcaggtcaggtcaggtcaggtc 1002
|:|||||
Db 8727 TTTTGACGGAGTCTCGTTCTGTCTACCCAGGCTGGAGTGCAGTGGCACAATCTCGGCTC 8668

QY 1003 attcagctccaccctcctcagttcaagc--ttctcctcctcagcgtcccaagtagctg 1060
|:|||||
Db 8667 ACTGCAACCTCCCGGTTCAAGCTATTCTCTCGCTCAGGCTCCCAAGTAGCTG 8608

QY 1061 ggattacaggctct 1074
|:|||||
Db 8607 GGACTACAGGCGCT 8594

RESULT 8
PCI-US95-07754A-4
: Sequence 4, Application PC/TUS9507754A
: GENERAL INFORMATION:
: APPLICANT: Baker, Barbara J
: APPLICANT: Whitham, Steven A
: TITLE OF INVENTION: Plant Virus Re
: NUMBER OF SEQUENCES: 6

LENGTH: 1143 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-04589-108

Query Match 3.7%; Score 122; DB 5; Length 1143;
Best Local Similarity 18.2%; Pred. No. 0.0039;
Matches 135; Conservative 106; Mismatches 200; Indels 300; Gaps 36;

QY 71 IYKNGDIVRFHRL-----KIQVKKETQG--ITSSGFASLTGEGTLGAPIIPRTSSKYFN 124
Db 335 LIEKNDIIEVTALPDHESIQLFQHAFGKEVPNENFEKLSLEVYVNYAKGLP----- 387
QY 125 FTTEHKKVZALRWAS-----THM-----SPSWTLKKL-----CDVQPMQ--- 160
Db 388 -----ALKVWGSLHLNRLTEWKSIAEHMKNNSYSGIIDKLSYDGLPKQOEM 437
QY 161 YFDLTCLGKAEVDGASFLKLVMD-----GTRTPFSNRVLQDLVLEGLDLSIHR 212
Db 438 FLDIACFLGEEK-----DYILOLESCHIGAEGYGLR-----ILDKSLVFISE 481
QY 213 LQNLITDILVYDNHVRVARSILK-VGSFLRIYSLHTKLSQMSNENQTMLSLEFHLHGTSY 271
Db 482 YNQVQMHDLIQDMGKIYVNFQKDGERSRLW-LAKEVEEVMNSNNTGTNAME--AIWVSSY 538
QY 272 GRGTRVLPESNSVDOLKDDLESANL-TANOHSV-----ICQEPDDSPSSGS 320
Db 539 SSTLRF---SNOAVKNMKR-LRVFNMGSRSTHYAIDYLPNNLRFCVCTNYPWESFPSTFE 594
QY 321 -----VSLYEVEVER 328
Db 595 LKMLVHLQLRHNSLRHLWTETKHLPSLRIDLSWSKRLTRPDTGMPNLYVNYLQCSN 654
QY 329 -----COOLSATILTDHOYLERTPLCAILK-----OKAPOQY-RIRA 364
Db 655 LEEVHSLGCKSKVIGLYLNDCKSLKRF-CVNVESLEYLGLRSCDSLEKLPYIGRMKP 713
QY 365 KLSYK-----PRRLFQSVKLHCPK-----CHL 387
Db 714 EIQIHMOGSGIRELPSIFQ-YKTHVTKLLWNMKNLVALPSSICRLKSLVLSVSGCSK 772
QY 388 LOEVPHE-GDLTIIFODGATKTPVVKLQNTSLYDSKIWTTRKQGRKVAVHVKNGILP 446
Db 773 LESLPEEIGDLNLRVFDASDTLILRPPSSIIRLNKL-IILMFRGKDGVF----- 823
QY 447 LSNBCLLLIEGGTILSEICKLSKNFNSVIPVRSGHEDLELDLSAPFLIOGTHHHYCKQC 506
Db 824 -----EFPVVAEGLHSLEYLNLSYCNLIDGGLP-----EEI 854
QY 507 SSLRSIONLSVYDRTSWIPSSVAEALGIVPLQYVFMFTLDDGTGVLEAY-LMDSKRF 565
Db 855 GSLSLKKLDSLRNFEHLPSISIAQ-----LQALQSLDKDCQRL 894
QY 566 FQIPASEVLMDLQKSDVHMDMFCPPGK---IDAYPMLECFIKSYNVT----- 613
Db 895 TOLPE-----LPPELNLHVDCMHMALK-FI-HYLVTRKKKLHRVK 932
QY 614 -NGTDNQICYQIFDTTVAEDV 633
Db 933 LDDAHNDMTYNLFAYTMFQNI 953

RESULT 5

US-08-261-663A-2
Sequence 2, Application US/08261663A
Patent No. 5571706

GENERAL INFORMATION:

APPLICANT: Barbara J
APPLICANT: Whitham, Steven A
TITLE OF INVENTION: Plant Virus Resistance Gene and Methods
NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:
ADDRESSEE: Margaret A. Connor, USDA-ARS
STREET: 800 Buchanan Street
CITY: Albany
STATE: CA
COUNTRY: USA
ZIP: 94710
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/261.663A
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Connor, Margaret A
REGISTRATION NUMBER: 30043
REFERENCE/DOCKET NUMBER: 0094.94
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 559-6067
TELEFAX: (510) 559-5777
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1144 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-261-663A-2

Query Match 3.7%; Score 122; DB 1; Length 1144;

Best Local Similarity 18.2%; Pred. No. 0.0039;
Matches 135; Conservative 106; Mismatches 200; Indels 300; Gaps 36;

QY 71 IYKNGDIVRFHRL-----KIQVKKETQG--ITSSGFASLTGEGTLGAPIIPRTSSKYFN 124
Db 335 LIEKNDIIEVTALPDHESIQLFQHAFGKEVPNENFEKLSLEVYVNYAKGLP----- 387
QY 125 FTTEHKKVZALRWAS-----THM-----SPSWTLKKL-----CDVQPMQ--- 160
Db 388 -----ALKVWGSLHLNRLTEWKSIAEHMKNNSYSGIIDKLSYDGLPKQOEM 437
QY 161 YFDLTCLGKAEVDGASFLKLVMD-----GTRTPFSNRVLQDLVLEGLDLSIHR 212
Db 438 FLDIACFLGEEK-----DYILOLESCHIGAEGYGLR-----ILDKSLVFISE 481
QY 213 LQNLITDILVYDNHVRVARSILK-VGSFLRIYSLHTKLSQMSNENQTMLSLEFHLHGTSY 271
Db 482 YNQVQMHDLIQDMGKIYVNFQKDGERSRLW-LAKEVEEVMNSNNTGTNAME--AIWVSSY 538
QY 272 GRGTRVLPESNSVDOLKDDLESANL-TANOHSV-----ICQEPDDSPSSGS 320
Db 539 SSTLRF---SNOAVKNMKR-LRVFNMGSRSTHYAIDYLPNNLRFCVCTNYPWESFPSTFE 594
QY 321 -----VSLYEVEVER 328
Db 595 LKMLVHLQLRHNSLRHLWTETKHLPSLRIDLSWSKRLTRPDTGMPNLYVNYLQCSN 654
QY 329 -----COOLSATILTDHOYLERTPLCAILK-----OKAPOQY-RIRA 364
Db 655 LEEVHSLGCKSKVIGLYLNDCKSLKRF-CVNVESLEYLGLRSCDSLEKLPYIGRMKP 713
QY 365 KLSYK-----PRRLFQSVKLHCPK-----CHL 387
Db 714 EIQIHMOGSGIRELPSIFQ-YKTHVTKLLWNMKNLVALPSSICRLKSLVLSVSGCSK 772
QY 388 LOEVPHE-GDLTIIFODGATKTPVVKLQNTSLYDSKIWTTRKQGRKVAVHVKNGILP 446
Db 773 LESLPEEIGDLNLRVFDASDTLILRPPSSIIRLNKL-IILMFRGKDGVF----- 823
QY 447 LSNBCLLLIEGGTILSEICKLSKNFNSVIPVRSGHEDLELDLSAPFLIOGTHHHYCKQC 506

Db 335 LIEKNDIIEVTALPDHESIOLEKFAHGFKEVNEFEKLSLEVYVYAKGLPL----- 387
Qy 125 FTEDHKVVEALRVAS-----THM---SPSWTLKL---CDVOPMO--- 160
Db 388 -----ALKVWGSLLHNLRLTEKSAIEHMKNSYSGIDKLSISYDGLPEKQOEM 437
Qy 161 YFDLTCLLGAEDVGFASFLKLVWD-----GTRTPFPSSRWVLIODLVLEGLDLSIHR 212
Db 438 FLDIACFLRGEK-----DYILOLESCHIGAEGYGLR-----ILDKSLVFISE 481
Qy 213 LONITIDILVYDNHVVHVARSLK-VGSFLRIYSLHTKLOSNNSENOTMLSLEFLHGGTSY 271
Db 482 YNOVOMHDLIOMGKYIVNFOKDPCERSRLW-LAKEVEEYMSNNTGTMAEIM--VSSY 538
Qy 272 GRGIRVLPESNSVDOLKDKLESANL-TANOHSV-----ICOSEPDSPSSSGS 320
Db 539 SSTLRP---SNOAVKNMKR-LRVFNMGSRSTHYAIDYLPNNLRFCVCTNYPWESFPSTPE 594
Qy 321 -----PRRLFOSVKLHCPK-----VSLYEYER 328
Db 595 LKMLVHLQRLHNSLRHLWTETKHLPSLRRLIDLSNKRTRTPDFTGMPNLEVYVNLQCSN 654
Qy 329 -----COOLSATILTDHOYLERTPLCALK-----OKAPOOY-RIRA 364
Db 655 LEEVHSLGCCSKVIGLYLNDCKSLKRF-CVNVESLEYLGLRSCDSLEKLEPEIYGRMKP 713
Qy 365 KLSRYK-----PRRLFOSVKLHCPK-----CHL 387
Db 714 EIOIHMOGSGIRELPSISFO-YKHTVKLLNMKNLVALPSSICRLKSLVSLVSGCSK 772
Qy 388 LOEVPHE-GDLDIIFODGATKTPVVKLQNTSLYDSKIWTTRKNQKRVAVHFKVNGILP 446
Db 773 LESLPEEGDLNLRVFDASDTLIRPPSSIIRLNKL-IILMFRGFKDGVHF----- 823
Qy 447 LSNECLLLIEGTLSEICKLSNKNFNSVIPVRSGHEDLELDSAPFLIOGTHHYGKQC 506
Db 824 -----EPPVAEGLSHLEYLNLSCYNLIDGGLP-----EEI 854
Qy 507 SSLRSIONLSLVDKTSWIPSSVAELGIVPLQYVFVMTFTLDDGTGVLEAY-LMDSKFX 565
Db 855 GSLSLKLRLDLSRNFEHLPSISAO-----LGAOSLDLKCQRL 894
Qy 566 FOIPASEVLDLDDLOKSDVMDMFCPPGK-----IDAYPWLECFIKSYNYT----- 613
Db 895 TOLPE-----LPPNELHVDCHMALK-FI-HYLVTKRKXLRHVY 932
Qy 614 -NGTDNOICQIIFDITVAEDV 633
Db 933 LDDAHNDTHNLFAITMFQNI 953

RESULT 2

US-08-310-912A-108
Sequence 108, Application US/08310912A
Patent No. 5981730

GENERAL INFORMATION:

APPLICANT: Ausubel, Frederick M.
APPLICANT: Staskawicz, Brian J.
APPLICANT: Brent, Andrew F.
APPLICANT: Dahlbeck, Douglas
APPLICANT: Katagiri, Fumitaki
APPLICANT: Kunkel, Barbara N.
APPLICANT: Mindrinos, Michael N.
APPLICANT: Yu, Guo-Liang

TITLE OF INVENTION: RPS2 GENE FAMILY, PRIMERS, PROBES, AND DETECTION
TITLE OF INVENTION: METHODS
NUMBER OF SEQUENCES: 208

CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.

STREET: 225 Franklin Street
CITY: Boston
STATE: MA

COUNTRY: USA
ZIP: 02110-2904
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
FILING DATE: September 22, 1994
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/227,360
FILING DATE: April 13, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Lech, Karen F.
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/254001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 100254
INFORMATION FOR SEQ ID NO: 108:
SEQUENCE CHARACTERISTICS:
LENGTH: 1143 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-310-912A-108

Query Match 3.7%; Score 122; DB 2; Length 1143;
Best Local Similarity 18.2%; Pred. No. 0.0039;
Matches 135; Conservative 106; Mismatches 200; Indels 300; Gaps 36;

Qy 71 IYKNGDIVRHL-----KIQYVKKEQO-----ITSSGFASLTFTGLGAPIIPTSKSYFN 124
Db 335 LIEKNDIIEVTALPDHESIOLEKFAHGFKEVNEFEKLSLEVYVYAKGLPL----- 387
Qy 125 FTEDHKVVEALRVAS-----THM---SPSWTLKL---CDVOPMO--- 160
Db 388 -----ALKVWGSLLHNLRLTEKSAIEHMKNSYSGIDKLSISYDGLPEKQOEM 437
Qy 161 YFDLTCLLGAEDVGFASFLKLVWD-----GTRTPFPSSRWVLIODLVLEGLDLSIHR 212
Db 438 FLDIACFLRGEK-----DYILOLESCHIGAEGYGLR-----ILDKSLVFISE 481
Qy 213 LONITIDILVYDNHVVHVARSLK-VGSFLRIYSLHTKLOSNNSENOTMLSLEFLHGGTSY 271
Db 482 YNOVOMHDLIOMGKYIVNFOKDPCERSRLW-LAKEVEEYMSNNTGTMAEIM--VSSY 538
Qy 272 GRGIRVLPESNSVDOLKDKLESANL-TANOHSV-----ICOSEPDSPSSSGS 320
Db 539 SSTLRP---SNOAVKNMKR-LRVFNMGSRSTHYAIDYLPNNLRFCVCTNYPWESFPSTPE 594
Qy 321 -----PRRLFOSVKLHCPK-----VSLYEYER 328
Db 595 LKMLVHLQRLHNSLRHLWTETKHLPSLRRLIDLSNKRTRTPDFTGMPNLEVYVNLQCSN 654
Qy 329 -----COOLSATILTDHOYLERTPLCALK-----OKAPOOY-RIRA 364
Db 655 LEEVHSLGCCSKVIGLYLNDCKSLKRF-CVNVESLEYLGLRSCDSLEKLEPEIYGRMKP 713
Qy 365 KLSRYK-----PRRLFOSVKLHCPK-----CHL 387
Db 714 EIOIHMOGSGIRELPSISFO-YKHTVKLLNMKNLVALPSSICRLKSLVSLVSGCSK 772
Qy 388 LOEVPHE-GDLDIIFODGATKTPVVKLQNTSLYDSKIWTTRKNQKRVAVHFKVNGILP 446
Db 773 LESLPEEGDLNLRVFDASDTLIRPPSSIIRLNKL-IILMFRGFKDGVHF----- 823
Qy 447 LSNECLLLIEGTLSEICKLSNKNFNSVIPVRSGHEDLELDSAPFLIOGTHHYGKQC 506


```

GN 2810458H16RIK.
OS MUS musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRIO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Akizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gotojori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant I.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio I.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.,
RA *Functional annotation of a full-length mouse cDNA collection.*;
RT Nature 409:685-690(2001).
DR EMBL; AK013364; BAB28810.1; -.
DR MGI; MGI:1920086; 2810458H16RIK.
SQ SEQUENCE 113 AA: 12740 MW: 08BBB9FC5D0C7088 CRC64;

```

Query Match 9.7%; Score 324; DB 11; Length 113;
 Best Local Similarity 73.3%; Pred. No. 3e-19; Indels 0; Gaps 0;
 Matches 63; Conservative 6; Mismatches 17;

```

QY 1 MSLVATNYIYPLNQLKGGTIVNYGVVKKFKPPYLSKGTGYDYSVVTIVDTNVLKTL 60
DB 1 MSAPVAPSTYIPLNQLKGGTIVNYGVVKKFKPPYLSKGTGYDYSVVTIVDTNVLKTCM 60
QY 61 LFSGNEALPIYKNGDIVRHLRI 86
DB 61 LFSRDYDTPSYKRVGDIVCFGLKV 86

```

```

RESULT 7
ID O13988 PRELIMINARY; PRT; 555 AA.
AC O13988;
DT 01-JAN-1999 (TrEMBLrel. 09, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE HYPOTHETICAL 64.1 KDA PROTEIN C26H5.06 IN CHROMOSOME 1.
GN SPAC26H5.06
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Oliver K., Harris D., Barrell B.G., Rajandream M.A., Wood V.;
RA Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RL EMBL; Z99126; CAB16192.2; -.
RW Hypothetical protein.
KW SEQUENCE 555 AA: 64111 MW: A79DAA95A0C4F803 CRC64;

```

Query Match 6.0%; Score 199.5; DB 3; Length 555;
 Best Local Similarity 23.1%; Pred. No. 6e-08;

```

Matches 109; Conservative 83; Mismatches 177; Indels 103; Gaps 23;
QY 18 KGTIVNYGVVKKFKPPYLS-KGT-DYCSVVTIVDT-----NVKLTCLLFSGNYEALPI 71
DB 18 KGTIVNYGVVKKFKPPYLS-KGT-DYCSVVTIVDT-----NVKLTCLLFSGNYEALPI 71
QY 38 KKNITLFGIYKVDTPSRQSLHGKDWYIVYLWDPCTDTSIGLQIHLFSKQGNLDPV 97
DB 38 KKNITLFGIYKVDTPSRQSLHGKDWYIVYLWDPCTDTSIGLQIHLFSKQGNLDPV 97
QY 72 IYKNGDIVRHLRIQVYKKTQGITSSGFA-----SLTFEGTLGAPIIPRTS----- 119
DB 72 IYKNGDIVRHLRIQVYKKTQGITSSGFA-----SLTFEGTLGAPIIPRTS----- 119
QY 98 IKQVGOPLLHQTILRSYRDRTOGLSKQDFRYALWPDFSSNSKDTLCPOPMPRLMKTGDK 157
DB 98 IKQVGOPLLHQTILRSYRDRTOGLSKQDFRYALWPDFSSNSKDTLCPOPMPRLMKTGDK 157
QY 120 -----SKYFNFTEDHKVYALRWAS-----THMS-PSWTLKLKCDVOPMQVFDLTC 166
DB 120 -----SKYFNFTEDHKVYALRWAS-----THMS-PSWTLKLKCDVOPMQVFDLTC 166
QY 158 EFOFALLLNKINDEQNRKNGELLSTSSARONQGLSPSVSFSLSQIIPHORCSFYA 217
DB 158 EFOFALLLNKINDEQNRKNGELLSTSSARONQGLSPSVSFSLSQIIPHORCSFYA 217
QY 167 QLLGKAEVDSGASFLKLVMDGTR-----TPFPS---NRVLIDLVLEDLSHIHRLQNL 216
DB 167 QLLGKAEVDSGASFLKLVMDGTR-----TPFPS---NRVLIDLVLEDLSHIHRLQNL 216
QY 218 OVI-KTWYDKNFTLVDTYENELFFPMSPYTSRRW-----GPGF-----RF 261
DB 218 OVI-KTWYDKNFTLVDTYENELFFPMSPYTSRRW-----GPGF-----RF 261
QY 217 TIDLIVDHHVARS-LKVGFLRIYLSHTKLNSENQTMLSLEFLHGTGTS--YGR 273
DB 217 TIDLIVDHHVARS-LKVGFLRIYLSHTKLNSENQTMLSLEFLHGTGTS--YGR 273
QY 262 SIRCILWDEHDFCRNYIKEGDYVVMKNVTRTKIDHLG-----YLECILHSDAKRYNM 314
DB 262 SIRCILWDEHDFCRNYIKEGDYVVMKNVTRTKIDHLG-----YLECILHSDAKRYNM 314
QY 274 GIRVLPESNDVQLK--KDLSEANLTANQHSVDVICS-----EPDSDFFSSGVSLSY 324
DB 274 GIRVLPESNDVQLK--KDLSEANLTANQHSVDVICS-----EPDSDFFSSGVSLSY 324
QY 315 SIEKVDEEPELNEIKSRRLYVN-----CONGIEAVIEKLSQSQSENPFIAH 364
DB 315 SIEKVDEEPELNEIKSRRLYVN-----CONGIEAVIEKLSQSQSENPFIAH 364
QY 325 EVER--COOLSATILDHQVLETPICAILKQK-----APQOYRIRAKLSYKPRLFOS 377
DB 325 EVER--COOLSATILDHQVLETPICAILKQK-----APQOYRIRAKLSYKPRLFOS 377
QY 365 ELKOTSNEYITAHVINEPASLKLTTITLHAPLQNLKPRKHLRVQVVDVFPKSLTQF 424
DB 365 ELKOTSNEYITAHVINEPASLKLTTITLHAPLQNLKPRKHLRVQVVDVFPKSLTQF 424
QY 378 VKLHCPKCH-----LLOEVPHEGDLIDIF--ODGATKTPVYVYKLTQNTSLYD 420
DB 378 VKLHCPKCH-----LLOEVPHEGDLIDIF--ODGATKTPVYVYKLTQNTSLYD 420
QY 425 AVLQSPSSSYVMFALLVRDYSNV-TLPIVFSDAAELNINSKIQPCNLAD 475
DB 425 AVLQSPSSSYVMFALLVRDYSNV-TLPIVFSDAAELNINSKIQPCNLAD 475

```

```

RESULT 8
Q9FNH7 PRELIMINARY; PRT; 463 AA.
ID O9FNH7
AC O9FNH7
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE GB|AAD29059.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=98069011; PubMed=9405937;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,062 bp covered by thirteen
RT physically assigned pl clones."
RL DNA Res. 4:291-300(1997).
DR EMBL; AB006700; BAB08953.1; -.
SQ SEQUENCE 463 AA: 53630 MW: A735DBF109155D7D CRC64;

```

Query Match 4.6%; Score 153.5; DB 10; Length 463;
 Best Local Similarity 22.8%; Pred. No. 0.00029;
 Matches 94; Conservative 59; Mismatches 184; Indels 95; Gaps 17;

```

QY 23 VNYGVWGAFFKPPYLSKGTGYDYSVVTIVD--QTNVKLTCLLFSGNYEALPIYKNGDIVR 80
DB 23 VNYGVWGAFFKPPYLSKGTGYDYSVVTIVD--QTNVKLTCLLFSGNYEALPIYKNGDIVR 80
QY 25 VNLIGVIVELG---FSGNSDCSLKIVDPWYSGSLPKFVARTIROLPRVESIGDIL 81
DB 25 VNLIGVIVELG---FSGNSDCSLKIVDPWYSGSLPKFVARTIROLPRVESIGDIL 81
QY 81 FHLRIQVYKKTQGI-----TSSGFASLTFFETLGAIIIPRTSSKYFNFTEDHKVYAL 136
DB 81 FHLRIQVYKKTQGI-----TSSGFASLTFFETLGAIIIPRTSSKYFNFTEDHKVYAL 136
QY 82 LSRVKIVLINRKITALCNETTTSSSFA--LFNGKHSDVSIPOSSPKFLMRQDNKFLSNL 139
DB 82 LSRVKIVLINRKITALCNETTTSSSFA--LFNGKHSDVSIPOSSPKFLMRQDNKFLSNL 139

```

RC TISSUE-Fetal brain;
 RX MEDLINE-99026138; PubMed-9806842;
 RA Fjiwara T., Saito A., Suzuki M., Shinomiya H., Suzuki T.,
 RA Takahashi E., Tanigami A., Ichihara A., Chung C.H., Nakamura Y.,
 RA Tanaka K.;
 RT "Identification and chromosomal assignment of USP1, a novel gene
 RL encoding a human ubiquitin-specific protease.";
 RL Genomics 54:155-158(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Testis;
 RX MEDLINE-21154917; PubMed-11230166;
 RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
 RA Ansorge W., Boecker M., Bloeker H., Bauersachs S., Blum H.,
 RA Lauber J., Duesterhoeft A., Beyer A., Koehler K., Strack N.,
 RA Newes H.W., Ottenwälder B., Obermaier B., Lampe J., Heubner D.,
 RA Wambutt R., Korn B., Klein M., Poustka A.;
 RT "Towards a catalog of human genes and proteins: sequencing and
 RT analysis of 500 novel complete protein coding human cDNAs.";
 RL Genome Res. 11:422-435(2001).
 CC -1- CATALYTIC ACTIVITY: UBIQUITIN C-TERMINAL THIOLESTER + H(2)O =
 CC UBIQUITIN + A THIOL.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C19; ALSO KNOWN AS
 CC FAMILY 2 OF UBIQUITIN CARBOXYL-TERMINAL HYDROLASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AB014458; BAA34703.1; -
 DR EMBL: AL117575; CAB55999.1; -
 DR EMBL: AL117503; CAB55967.1; -
 DR MEROPS: C19.019; -
 DR MIM: 603478; -
 DR InterPro: IPR001394; UCH-2.
 DR Pfam: PF00442; UCH-1; 1.
 DR Pfam: PF00443; UCH-2; 1.
 DR PROSITE: PS00972; UCH_2_1; 1.
 DR PROSITE: PS00973; UCH_2_2; 1.
 DR PROSITE: PS02335; UCH_2_3; 1.
 DR Ublquitin conjugation; Hydrolase; Multigene family.
 KW ACT_SITE 90 90 BY SIMILARITY.
 FT ACT_SITE 584 584 BY SIMILARITY.
 FT ACT_SITE 593 593 BY SIMILARITY.
 FT ACT_SITE 593 593 BY SIMILARITY.
 SQ SEQUENCE 785 AA; 88207 MW; 50AA2817A60810AF CRC64;

Query Match 3.1% Score 104; DB 1; Length 785;
 Best Local Similarity 19.8% Pred. No. 5.9;
 Matches 96; Conservative 64; Mismatches 213; Indels 112; Gaps 18;

QY 154 CDVQPMQVFDLTC---QLLGRAEVGASFLKLVGDGTRTPFPSSNRVLQDL-----VLEG 205
 DB 135 KEDSLASYELICSLSQSLIIISVEQLASFLNPKYKIDELATQPRRLNTLRELNPMYEG 194
 QY 206 DLSHTHRLQNLITDIL--VYDNHVRHVARSLKVGSLFIYSLHTKLOSNSSENQ;MLSLEF 263
 DB 195 YLQHDQAQ-----EVLCILGNIQETCOLLAKKEVKNVAELPKVEIHPHKEEM----- 243
 QY 264 HLHGTSYGRGIRVLPSNSVDVQLKDDLESANLTANOHSVICOSEPDOSFSGSVSL 323
 DB 244 -----NGINSIEMDSMRHSEDFEKLPGKNGKFKSDTEF 277
 QY 324 YEVEPCQOLSATILTDHOYLERTPLCALLKOKAQOQYIRAKLSYKRRPRLFQSVKLHCP 383
 DB 278 GNMKKKVKLS-----KEHOSLEENORQTRSKRKATSD-----TLESPPKTIIP 319
 QY 384 KCHLLQEVPH-----EGDLDIIIFODGATKTPVVKLQNTSLYDSKIWTITKNOKGRVAVHFV 439

DB 320 KYISENPRSPQSKSRKINWLKATQPSILLSKFCSL--GKITNQGVKGQS----- 371
 QY 440 KNGILP-----LSNECLLLIEGGTSLSEICKLSNKFNSVIVRSGHEDL--ELL 486
 DB 372 KENECDPEEDLGKCESDNTTNGCGLESPGNTVTPV-----NVNEVKPKNGEEOIGFELY 426
 QY 487 DLSAPELIQG-TIHHYGCKOCSSL-----RSTONLSLVLDKTSWTPSPVAAEALGIVPLQYV 541
 DB 427 EK-----LFQGLVLRTRCLECESLTERREDFQDISVPVQDEL--SKVESSEISPEPKT 480
 QY 542 FVMTFTLDDGTGVLEAYLMSDKFF-----QIPASEVLMDDDLQKSVDMIMDMFCPPGI 595
 DB 481 EMKTLWAIISOFAFASVERIVGDKYFCNCCHHYTEAERSLLFDKMPKPEVITTHLKCFASAGL 540
 QY 59% KIDAY 500
 DB 541 EFDY 545

RESULT 14
 UVRA_CHLMU STANDARD: PRT; 1787 AA.
 AC Q9PK60;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DE EXCINUCLEASE ABC SUBUNIT A.
 GN UVRA OR TC0610.
 OS Chlamydia muridarum.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83560;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MoPn / Nigg.
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Elsen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
 RT pneumoniae Ar39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 CC -1- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
 CC CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
 CC PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UVRA IS AN ATPASE
 CC AND A DNA-BINDING PROTEIN THAT PREFERENTIALLY BINDS SINGLE-
 CC STRANDED OR UV-IRRADIATED DOUBLE-STRANDED DNA (BY SIMILARITY).
 CC -1- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVRA, UVRB AND UVRC.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
 CC (ABC TRANSPORTERS). CONTAINS FOUR ABC DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE002329; AAF39441.1; -
 DR TIGR: TC0610; -
 DR InterPro: IPR003439; ABC_transportr.
 DR Pfam: PF00005; ABC_tran; 2.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 KW SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
 KW DNA-binding; Zinc-finger; Complete proteome.
 FT NP_BIND 32 39 ATP (POTENTIAL).
 FT NP_BIND 625 632 ATP (POTENTIAL).
 FT NP_BIND 964 971 ATP (POTENTIAL).
 FT NP_BIND 1509 1516 ATP (POTENTIAL).
 FT ZN_FING 719 742 C4-TYPE.

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; J03358; AAA61190.1; -;
DR PIR; A31943; TVHUFE.
DR HSP; P11362; 1FGI.
DR MW; 176942; -;
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001060; FCH.
DR InterPro; IPR000980; SH2.
DR InterPro; IPR001245; Tyr_kin.
DR Pfam; PF00611; FCH; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00017; SH2; 1.
DR PRINTS; PR00109; TYRKINASE.
DR SMART; SM00055; FCH; 1.
DR SMART; SM00352; SH2; 1.
DR SMART; SM00219; TYRK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS50001; SH2; 1.
DR Transferase; Tyrosine-protein kinase; Proto-oncogene; ATP-binding;
KW Phosphorylation; Nuclear protein; SH2 domain; Polymorphism.
FT DOMAIN 460 550
FT NP_BIND 563 816 PROTEIN KINASE.
FT BINDING 569 577 ATP (BY SIMILARITY).
FT ACT_SITE 591 591 ATP (BY SIMILARITY).
FT MOD_RES 684 684 BY SIMILARITY.
FT VARIANT 714 714 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT V -> L.
FT /FTID=VAR_006282.
SQ SEQUENCE 822 AA: 94623 MW: 0491CD6392DB415 CRC64;

Query Match 3.28; Score 107; DB 1; Length 822;
Best Local Similarity 20.08; Pred. No. 3.8;
Matches 127; Conservative 82; Mismatches 213; Indels 214; Gaps 29;

OY 37 LSKGTDYCVYTIYDQTNVLTCLLFGSGNEALPIYK-----NGDIVRFHRLKIQVY 89
DB 49 LCNQVDESTVOMYVSNVSKSWLLMIQOTQLSRIMKTHAEDLNSGFL---HRLTMMI- 104
OY 90 KETQGITSSGFSALTEGTLGAPIIPTSS-----KYFNFTEDHKVVEAL 136
DB 105 -KDKQVKKK---YGVHQQTAEAMIKVTKTELEKLCYSRQLIKEMNSAKERYK--EAL 158
OY 137 RWAST-----HMSPSWTLKL-CDVOPMQYFDLTCLLGKAEVDGASFL 180
DB 159 AKGTEKAKERYDKATMKLHMLNQYVLAUKGALQHQNYDITLPLLLDSLOKMOEM 218
OY 181 LKWNVDGTRTPFSSRVRLIQDLVLEGLSHIRLQNLITDIL-----YVDNHVHVARSLKVG 236
DB 219 IKALKGIFDEYSQ-----ITSLVTE-EIVNVHKEIQMSVEQIDPSTEXNFFIDVHRTTA-- 271
OY 237 SFLRIYSLHTKLQ-----SNSENQTMLSLEFHLHGGTSYGRGI-----RV 277
DB 272 -----AKEQETFEFTSLLEENENIQANEIMWNLTAEISLOVMUKTLAEELMOTOOM 322
OY 278 LPESNSVDYDLKDKLESANLTANQSHSDVICOSEPDGFPSSGVSLSVEVERCOOLSATIL 337
DB 322 LNKKEAVLELEKRIEESSECEKKSQIV-----LLL 354
OY 338 TDHLYERTPLCALIKAKAPQOYRIKRLSKYKPRFLQSVKVLKCPKCHLLQEPHGEDL 397
DB 355 SOKQALEE-----LKQSVQOOLRTEAKFSAQK-----ELLEQKVOE--- 390
OY 398 DIIFDQGATKTPVVKLQNTSLYDSKTIWTKMKQGRKVAHVFKNN--GIL--PLSNECLL 453

DB 391 ----NDGKEPPPVVYEE-----DARSVTSMERKLERLSKFESIRHSTAGIIRSPKS----- 437
OY 454 LIEGGTLSEICKLSNK-----FNSVIVPRSGHEDLELLDLSAPFLI----- 494
DB 438 AVGSALSMDISISEKPLAEODWYHGAIP---RIEAQELLKKQGDPLVRESHGKPGEYVL 494
OY 495 -----OGT-----IHGYCKOCSSLR--IQNLNSLVDT 522
DB 495 SVYSDGORRHFIQYVDNMVRFEGTGFNSIPQLIDHHYTKQVITKKSGVLLNPIPKDK 554
OY 523 SWIPSS-----VAEALGIVPLQYVYVMTFTLDDGTGV 554
DB 555 KWILSHEDVILGELLG--KGNFGEVYKGTLDKDKTSV 588
RESULT 11
ID SP2E_BACME STANDARD; PRT; 585 AA.
AC P49600;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE STAGE II SPOULATION PROTEIN E (EC 3.1.3.16) (FRAGMENT).
GN SPOIIE.
OS Bacillus megaterium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1404;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=VT1660;
RX MEDLINE=96249699; PubMed=8830262;
RA Barak I., Behari J., Olmedo G., Guzman P., Brown D.P., Castro E.,
Walker D., Westpheling J., Youngman P.;
RT Structure and function of the Bacillus Spolie protein and its
RT localization to sites of sporulation septum assembly.";
RL Mol. Microbiol. 19:1047-1060(1996).
CC -!- FUNCTION: NORMALLY NEEDED FOR PRO-SIGMA E PROCESSING DURING
CC SPOULATION BUT CAN BE BYPASSED IN VEGETATIVE CELLS. ACTIVATES
CC SPOIIAA BY DEPHOSPHORYLATION (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: A PHOSPHOPROTEIN + H(2)O = A PROTEIN +
CC ORTHOPHOSPHATE (THIS ENZYME IS SERINE/THREONINE SPECIFIC).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. POLAR SEPTUM (BY
CC SIMILARITY).
CC -!- SIMILARITY: CONTAINS A PP2C-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; U26836; AAB58072.1; -;
DR InterPro; IPR003589; PP2C_catalytic.
DR InterPro; IPR003588; PP2C_sig.
DR SMART; SM00332; PP2CC; 1.
DR SMART; SM00331; PP2C_SIG; 1.
KW Sporulation; Hydrolase; Transmembrane.
FT NON_TER 1 1
FT TRANSHEM 40 57 POTENTIAL.
FT TRANSHEM 70 86 POTENTIAL.
FT DOMAIN 366 585 PP2C-LIKE.
SQ SEQUENCE 585 AA: 65690 MW: 90E9ACF1D3E21D01 CRC64;

Query Match 3.28; Score 105; DB 1; Length 585;

Best Local Similarity 19.48; Pred. No. 3.2;
Matches 120; Conservative 89; Mismatches 202; Indels 208; Gaps 29;
OY 46 VVT--IVDOTNW-----KLTCLLFGSGNEALPIYKNGDIVRFH-----RLKIQVYKKT 93


```

Db 368 TOF--YITRVPDVKEVYSYDRKTKRPSHKGAGKGENIFOVLVKDASTOLANN 425
QY 341 QY 342
Db 426 TY 427

RESULT 5
EX02.BPT4
ID EX02.BPT4 STANDARD; PRT; 560 AA.
AC P04522;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EXONUCLEASE SUBUNIT 2 (EC 3.1.11.-) (PROTEIN GP46).
GN 46.
OS Bacteriophage T4.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae;
OC T4-like phages.
OX NCBI_TaxID=10665;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85257446; PubMed=4018026;
RA Gram H., Rueger W.;
RT "Genes 55, alpha gt, 47 and 46 of bacteriophage T4: the genomic
RT organization as deduced by sequence analysis.";
RL EMBO J. 4:257-264 (1985).
RN [2]
RP SEQUENCE FROM N.A.
RA Kutter E., Arisaka F., Kunisawa T., Tsugita A., Mosig G.,
RA Mesyanzhinov V., Ruger W., Stidham T., Thomas E.;
RT "Bacteriophage T4 genome analysis.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 478-560 FROM N.A.
RA Hsu T., Karam J.;
RL Submitted (APR-1987) to the EMBL/GenBank/DBJ databases.
CC 1- FUNCTION: EXONUCLEASE INVOLVED IN PHAGE DNA RECOMBINATION,
CC REPLICATION, AND REPAIR.
CC 1- SUBUNIT: CONSISTS OF TWO SUBUNITS: GP47 AND GP46.
CC 1- SIMILARITY: STRONG TO T5 PROTEIN D13 AND TO YEAST RAD52.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X01804; CAA25945.1;
CC EMBL; M15080; AAA32516.1;
CC EMBL; AF158101; AAD42472.1;
CC EMBL; M10160; AAC05392.1;
CC PIR; A04298; NCBPX6.
CC NP_BIND 36 43 ATP (POTENTIAL).
SQ SEQUENCE 560 AA; 63613 MW; D107829431BAB3FC CRC64;

Query Match 3.7%; Score 122; DB 1; Length 560;
Best Local Similarity 18.7%; Pred. No. 0.16;
Matches 102; Conservative 78; Mismatches 168; Indels 198; Gaps 22;

QY 65 NYEALPIYKNGDIFVRHLKIOYKKTGKIT-----SSGFASLTFTGTLGAPLIPR 117
Db 3 NFKLRVKYKINMSYGVGNGIDQLDKVOKTLITGRNGGKSTMLEAITP-GLFGKRPDRV 61

QY 118 TSSRYFNFTTDEHKMVEALRWASTHMSPTLLKLCDDVQPMQY-----F 152
Db 62 KKGQLINSTNKKELLVE--LW-----MEYDEKKYIKRGQKPNVF 99

QY 163 DLTC-----QLLCKAEVDGASFLKLVWDGTR--TPF-----PSWRV 196

```

```

Db 100 EITVNGTRLNESASSKDFQAEFEQLIG---MSYASFQKQIVLGLTAGYTPFMGLSTPARRK 156
QY 197 LIQDLVLEGLDLSHIHRLQNLITIDILVYDNHVVHVARSLKVGSEFLRYSLSLTKLQ-SMNSEN 255
Db 157 LVEDLLEVTGLAEMDKLNKALIRELSNQVLDVKKDSIIQOIKIYNDNVERQKLTGDN 216
QY 256 QTMLSLEFHLHGTSYGRIRVLPSNSDVDDKKDLKSANUTANQHSVDVICQSEPDPSF 315
Db 217 LTRLQ-----NMYDDLAKERTLKSEIEAN---ERLVNVLVDEDTDAF 258
QY 316 PSSGS-----VSLY-----EVERC-----OO 331
Db 259 NKIGQEAFLIKSKIDSYNKVINNYHEGGLCTCLSQLSGDKVKVSKDKVSECTHSFEQ 318
QY 332 LSA-----TILTDHQLERTPLCAILK-----OKAPOQYRIR 363
Db 319 LSTHRDLNLKVLVDEYRDNIKTQOSLANDIRNKKQSLIAAVDKAKKAAIEKASSEFIDH 378
QY 364 AKLRSYKPRRLFSQVKLHCPCCHLLOEVPHEGDLDIIFQDGATKTPVVKLQNTSLYDSKI 423
Db 379 ADEIALLOEELDKIVK---TKTNLVMEKYHVRGILTDMLKDSGIGAIK-KYIPLFNKOI 434
QY 424 WTKKQGRKRVAVHFVKNGNGLPLSNCELLIEGGTSLSEICKLSNKFNSVIPRSGHEDL 483
Db 435 -----NHYLKIMEA---DIVFTLDEEFNETIKSR-GREDF 465
QY 484 ELLDLS 489
Db 466 SYASFS 471

RESULT 6
GATL_YEAST STANDARD; PRT; 510 AA.
AC P43574;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRANSCRIPTIONAL REGULATORY PROTEIN GATL.
GN GATL OR YFL021W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GRF88;
RX MEDLINE=96182087; PubMed=8622886;
RA Coffman J.A., Rai R., Cunningham T., Svetlov V., Cooper T.G.;
RT "Gatlp, a GATA family protein whose production is sensitive to
RT nitrogen catabolite repression, participates in transcriptional
RT activation of nitrogen-catabolic genes in Saccharomyces cerevisiae.";
RL Mol. Cell. Biol. 16:847-858 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=95400292; PubMed=7670463;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M., Tashiro H., Eki T.;
RT Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae.";
RL Nat. Genet. 10:261-268 (1995).
CC 1- FUNCTION: POSITIVE REGULATOR OF MULTIPLE NITROGEN CATABOLIC GENES.
CC 1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC 1- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC

```

Db 69 VNVFFSQNFEDLPFIORVGDIYVRHARLQHYNDARLQNVNMYRSMCLFICNDKEAP 128
 QY 114 IIPRT-----SSRYE-----NFTEDH--KMVEALRVWASTHMSPTL--LKLC 154
 Db 129 LEFVNEEDTNNYFSTPNFSGKSTQGHETKILKDLKWSKDYFSNNDVVEQVKA 168
 QY 155 DVOPMOY-----FDLTCLLGAKEVDSGLKLVMDGT-----RTFFPSWEVLIDQL 201
 Db 189 DIETAMKNTDFOLLAKVTEISNDQVNTVSLNDSTGQTWTGHLFRKRFPH----- 240
 QY 202 VLEGDLSHIHLQNLITDILVYDNHVVARSKVGSLFRLIYSHIKLOSNNSEN 255
 Db 241 LVKGDVLRKSAKEDNSLIFSSHNLK-----FFSFSSIHKKLKSSISSU 288

RESULT 2
 ID TEBH_EUPCR STANDARD; PRT; 460 AA.
 AC Q06183; 1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DE 01-JUN-1994 (Rel. 29, Last annotation update)
 DT TELOMERE-BINDING PROTEIN HOMOLOG.
 OS Euplotides crassus.
 OC Eukaryota; Alveolata; Ciliophora; hypotrichs; Euplotida; Euplotidae;
 OC Euplotides.
 OX NCBI_TaxID=5936;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93126105; PubMed=1480483;
 RA Wang W., Skopp R., Scofield M., Price C.;
 RT "Euplotides crassus has genes encoding telomere-binding proteins and
 telomere-binding protein homologs."
 RL Nucleic Acids Res. 20:6621-6629(1992).
 CC -1- FUNCTION: MAY BIND TELOMERIC TAG4 SEQUENCES.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -1- SIMILARITY: BELONGS TO THE TELOMERE-BINDING PROTEIN ALPHA SUBUNIT
 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EXBL; M96819; AAA29128.1; -
 DR PIR; S35525; S35525.
 DR InterPro; IPR003415; Telo_bind_alpha.
 DR Pfam; PF02307; Telo_bind_alpha; 1.
 KW DNA-binding; Nuclear protein; Telomere; Multigene family.
 SK SEQUENCE 460 AA; 53360 MW; ED1C141385A0B5FE CRC64;

Query Match 4.98; Score 163.5; DB 1; Length 450;
 Best Local Similarity 21.66; Pred. No. 0.0001.
 Matches 117; Conservative 74; Mismatches 195; Indels 155; Gaps 24;

QY 9 YIYTPNLQ--KGGTIVNVYGVYFFKPPYLSKGTGYCSVYTIIVDOT-NVK-----LTC 59
 Db 19 YEYTEIGSIEENEASINFYAVVIDACEFYKVKDEKKYCYLKVITDTHNVKRGDDNFAIV 78
 QY 60 LIFSGNTEALPIIYKNGDIDVFRHLKIOVYKKEITGII-----SSGFASLIFEGLGAPI 114
 Db 79 ALQSRKFDELPIIQCGDIIIRVHRAEYN-YKDDQHYFKLNYSYSSWALFSADEEVAPEV 137
 QY 115 IP-----RTSSKYVFTTEDHKMVEALRVWASTHMSPTSLKLCQVCPMOYFD 163
 Db 138 IKDEGDDFTYRSYAYSGKQYNEFDQDQKLLNTRANNKSYFAKN-----DV----- 183
 QY 164 LTCQLLGRAEVDGSAFLAKVMDGTPTFPFSWRVLIDLVLEGDLSHIHLQNLITDILVY 223

Db 184 -----IIDEMYTPLSQROEBGDFNVGVKVTQIVHRDYITSDLRVK 224
 QY 224 DN-----HVHVARSLKVGSLFRIY-SLHTKLOSMSSENQTMLSLEPHLHGTSYGRGIRV 277
 Db 225 DISKATWELTVSRR-----KFRPLYEGVLIKIRSVNIDSE-----TERERCLEL 268
 QY 278 LPESN-----SDVDLKKDLKLESANLTANOHS-----VICOSEPDDSFSPSGSVSLYEVR 328
 Db 269 AHSNTIMTFVPSRLAKSLDSQISLSPDKVELIKKIVLTPE----- 311
 QY 329 COOLSATILTDHOYLERTPLCALIKOKAPQ--YIRAKLRYSYKPRRLQFOSVKLHCPKCH 386
 Db 312 --VLAITTFGDIYSELPLTELSEIFEDVTDKDAVFRARFSLKITPDREDVEYVEETPK-- 367
 QY 387 LLOEVPHEGDLDIIFODGATKT--PVVKLO-----NTSLYDS--KIWT-TKNOKGRKAV 436
 Db 368 -----GAPRSKPYKVQFLIKDPSTALNDNLKYIYLSHGDGKGEFF- 409
 QY 437 HFVKNNGILPLSLNCLLIEGTLSSEICKLSNKFNSVIVRSRSGHEDLELDLSAPFLIOG 496
 Db 410 -----PGVDPSAQ-----TPSGHSLK-RKYASTLMKFNHIDAVLEKVGGAFFIRD 455
 QY 497 T 497
 Db 456 T 456

RESULT 3
 ID TEBA_OXYNO STANDARD; PRT; 495 AA.
 AC P29549;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE TELOMERE-BINDING PROTEIN ALPHA SUBUNIT (TELOMERE-BINDING PROTEIN 56
 DE KDA SUBUNIT) (TEBP ALPHA).
 GN MAC-56A AND MAC-56K AND MAC-56S.
 OS Oxytricha nova.
 OC Eukaryota; Alveolata; Ciliophora; hypotrichs; Stichotrichida;
 OC Oxytrichidae; Oxytricha.
 OX NCBI_TaxID=5945;
 RN [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RX MEDLINE=92035001; PubMed=1840510;
 RA Gray J.T., Celandier D.W., Price C.M., Cech T.R.;
 RT "Cloning and expression of genes for the Oxytricha telomere-binding
 RT protein: specific subunit interactions in the telomeric complex."
 RL Cell 67:807-814(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92249771; PubMed=1577273;
 RA Mitcham J.L., Lynn A.J., Prescott D.M.;
 RT "Analysis of a scrambled gene: the gene encoding
 RT alpha-telomere-binding protein in Oxytricha nova."
 RL Genes Dev. 6:788-800(1992).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
 RX MEDLINE=99091054; PubMed=9875850;
 RA Horvath M.P., Schweiker V.L., Bevilacqua J.M., Ruggles J.A.,
 RA Schultz S.C.;
 RT "Crystal structure of the Oxytricha nova telomere end binding protein
 RT complexed with single strand DNA."
 RL Cell 95:963-974(1998).
 CC -1- FUNCTION: MAY FUNCTION AS PROTECTIVE CAPPING OF THE SINGLE-
 CC STRANDED TELOMERIC OVERHANG. MAY ALSO PARTICIPATE IN TELOMERE
 CC LENGTH REGULATION DURING DNA REPLICATION. BIND SPECIFICALLY TO THE
 CC TAG4-CONTAINING EXTENSION ON THE 3' STRAND AND PROTECT THIS REGION
 CC OF THE TELOMERE FROM NUCLEASE DIGESTION AND CHEMICAL MODIFICATION.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- MISCELLANEOUS: THE SEQUENCE OF THE A (OR ALANINE) VERSION IS
 CC SHOWN. THE S (OR SERINE) VERSION DIFFERS IN ONLY TWO POSITIONS,
 CC AND THE K VERSION IN ONLY ONE. THERE MAY BE OTHER VERSIONS IN

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification; the primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB952446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 568 BP; 167 A; 115 C; 96 G; 178 T; 12 other;

Query Match 22.0%; Score 419; DB 22; Length 568;

Best Local Similarity 93.2%; Pred. No. 6.1e-106;

Matches 465; Conservative 0; Mismatches 30; Indels 4; Gaps 3;

QY 1411 aatagtgaattctctgtgagatcgccacgaag--acctggaacttttggacctttcag 1468

DB 541 ANAAGGGAATTCGGGGAATNGGCCAAGGAGGACCTGGGAATTTGGACCTTTTCAG 482

QY 1469 caccattctttacagaagaacaatacatcactatggtatgaacagtggttctagtga 1528

DB 481 CACCATTTTATACAGGGAACAAAACATCACTAGGATGTAACAGTGTNTAGTTGA 422

QY 1529 gatccatacaaatctaaatt--ccctggttgataaaacatcggtgattccttc-ttctgt 1586

DB 421 GATCCATACAAAATNTAAATTCCTCGTGTGATAAACAICGTGGATTCNTTCTCTGT 362

QY 1587 ggcagaagcactgggtattgtaccctccaatattgtttgttatgacctttacacttga 1646

DB 361 GGCAGAAGCACTGGGTATTGNACCCCTCCAATATGTTGTATGACCTTTACACTTGA 302

QY 1647 tgatgaacagagactactagaagcctatctctatgattcttgacaaaattcttcagattcc 1706

DB 301 TGATGAACAGAGTACTAGAAAGCCTATNTCATGATTNTGACAAATNTTCCAGATTCC 242

QY 1707 agcatcagaagttctgagatgagcttcacagaaagtgtgatatgatatgatatgatat 1766

DB 241 ACCATCAGAAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 182

QY 1767 gttttgtctccaggaataaaaattgatgcataatccgtggttggaatgcttcatcaagtc 1826

DB 181 GTTTTGTCTCCAGGAATAAAAAATTGATGCATATCCGTGGTTGGAATGCTTCATCAAGTC 122

QY 1827 atacaatgtcacaatgaacagataatacaatttgcattatcagatttttgaccacagtc 1886

DB 121 ATACAATGTCAAAATGGAACAGATAAATCAAAATTTGCTATCAGATTTTGTACACACAGT 62

QY 1837 tgcagaagatgtaattctaa 1905

DB 61 TGCAGAAGATGTAATCTAA 43

RESULT 9

AAH05347

ID AAB05347 standard; cDNA; 866 BP.

XX AC AAH05347;

XX 26-JUN-2001 (first entry)

XX Human cDNA clone (5'-primer) SEQ ID NO:2182.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX

XX

XX

XX Homo sapiens.

OS EP1074617-A2.

PN 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI: 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs

XX Claim 1: SEQ ID 2182; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.

XX Sequence 866 BP; 254 A; 162 C; 171 G; 274 T; 5 other;

QY Query Match 18.0%; Score 342; DB 22; Length 866;

Best Local Similarity 91.2%; Pred. No. 1.4e-84;

Matches 384; Conservative 0; Mismatches 30; Indels 7; Gaps 2;

QY 1 atgtcttgggtccagcaaaaattatatatatcacccctggaatcaacttaagggtggt 60

DB 439 atgtcttgggtccagcaaaaattatatatatcacccctggaatcaacttaagggtggt 498

QY 61 acaattgtcaatgtctatgtgtgtgaagttctttaagcccccattatcaagcaaaagga 120

DB 499 acaattgtcaatgtctatgtgtgtgaagttctttaagcccccattatcaagcaaaagga 558

QY 121 actgattattgtcagttgttaactattgtggaccagacaaatgaaacttaacttgcctg 180

DB 559 actgattattgtcagttgttaactattgtggaccagacaaatgaaacttaacttgcctg 618

QY 181 ctctttagtgaaactgaagccctcccaataattataaaaatggagatattgttcgc 240

XX

```
Oy 1580 cttctgtgacagaacactgggtattgtaccctcccaatatgtgtttgttatgaccttta 1639
Dy 601 cttctgtgacagaacactgggtattgtaccctcccaatatgtgtttgttatgaccttta 660
Oy 1640 caattgatgatgaacagagactacagaagcctatctcatgattctgaacaaattcttcc 1699
Dy 661 caattgatgatgaacagagactacagaagcctatctcatgattctgaacaaattcttcc 720
Oy 1700 agattccagatcagaagtctctgattgagatgacattcagaaaagtgtggatgatca 1759
Dy 721 agattccagatcagaagtctctgattgagatgacattcagaaaagtgtggatgatca 780
Oy 1760 tggatattgttctcctcagagaataaaaaattgatgcataatcgtggtggaatcttca 1819
Dy 781 tggatattgttctcctcagagaataaaaaattgatgcataatcgtggtggaatcttca 840
Oy 1820 tcaagtcatcaaatgtccaaatggaacagataatcaaatgttctcatcagatttttgaca 1879
Dy 841 tcaagtcatcaaatgtccaaatggaacagataatcaaatgttctcatcagatttttgaca 900
Oy 1880 ccacagttcagaagaatgtaattctaa 1905
Dy 901 ccacagttcagaagaatgtaattctaa 926

RESULT 6
AAH05303
ID AAH05303 standard; cDNA: 576 BP.
AC AAH05303;
DT 26-JUN-2001 (first entry)
XX Human cDNA clone (5'-primer) SEQ ID NO:2138.
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
KW Homo sapiens.
OS Homo sapiens.
XX EP1074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 1; SEQ ID 2138; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
```

```
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 576 BP; 167 A; 113 C; 114 G; 173 T; 9 other;

Query Match 28.6%; Score 545; DB 22; Length 576;
Best Local Similarity 98.7%; Pred. No. 9.1e-141;
Matches 545; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy 1 atgtcttgggtccagcaacaatttatatatatacacccctgaatcaacttaaggtggt 60
Dy 25 atgtcttgggtccagcaacaatttatatatatacacccctgaatcaacttaaggtggt 84
Oy 61 acaattgtcaatgtctatgtgtgtgaagtcttttaagccccccatataatgaagca 120
Dy 85 acaattgtcaatgtctatgtgtgtgaagtcttttaagccccccatataatgaagca 144
Oy 121 actgattattgtcagttgtaactattgtgaccagacaaaatgtaaaactaacttgcctg 180
Dy 145 actgattattgtcagttgtaactattgtgaccagacaaaatgtaaaactaacttgcctg 204
Oy 181 ctcttagtggaactatgaagcccttcccaataattataaaatggagatatgttcgc 240
Dy 205 ctcttagtggaactatgaagcccttcccaataattataaaatggagatatgttcgc 264
Oy 241 ttccacaggtgaagattcaagtataataaaaggagaccaggtatcacagctctgac 300
Dy 265 ttccacaggtgaagattcaagtataataaaaggagaccaggtatcacagctctgac 324
Oy 301 ttgcatctttgacgtttgagggaaactttggagccccctatcatcacctgcacttcaagc 360
Dy 325 ttgcatctttgacgtttgagggaaactttggagccccctatcatcacctgcacttcaagc 384
Oy 361 aagatttttaacttcaactactgagaccacaaaatgtagaagccttaactgtttgggca 420
Dy 385 aagatttttaacttcaactactgagaccacaaaatgtagaagccttaactgtttgggca 444
Oy 421 tctactcatatgtcacccgtcttggacattactactaaaattgtgtgattcagccaatgcag 480
Dy 445 tctactcatatgtcacccgtcttggacattactactaaaattgtgtgattcagccaatgcag 504
Oy 481 tattttgacctgactgtcagctcttgggcaagcaagagtgagcagagcatcattctt 540
Dy 505 tattttgacctgactgtcagctcttgggcaagcaagagtgagcagagcatcattctt 564
Oy 541 cttaaagggtatgg 552
Dy 565 cttaaagggtatgg 576

RESULT 7
AAH07649
ID AAH07649 standard; cDNA: 854 BP.
XX
AC AAH07649;
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (5'-primer) SEQ ID NO:4484.
XX
```

PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0523317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-7G-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR P-PSDB; AAM40125.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1: SEQ ID NO 1484; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 3622 BP; 1143 A; 656 C; 657 G; 1166 T; 0 other:
XX
Query Match 99.8%; Score 1901.8; DB 22: Length 3622;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1903; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 atgtcttgggtccagcaacaattatataataacacccctgaatcaacttaagggtgt 60
DB 131 atgtcttgggtccagcaacaattatataataacacccctgaatcaacttaagggtgt 190
QY 61 acaattgtcaattgtctatggtgtgtggaagttctttaaagcccccatacttaagcaaaaga 120
DB 191 acaattgtcaattgtctatggtgtgtggaagttctttaaagcccccatacttaagcaaaaga 250
QY 121 actgattattgtcagttgtaactattgtggaccagacaaaatgtataaacttaactgtcctgc 180
DB 251 actgattattgtcagttgtaactattgtggaccagacaaaatgtataaacttaactgtcctgc 310
QY 181 ctctttagtgaactgaagccctcccaataattataaaatggagataattgttcgc 240
DB 311 ctctttagtgaactgaagccctcccaataattataaaatggagataattgttcgc 370
QY 241 ttccacaggctgaagattccaagtatataaaaaaggagactcagggtatccaccgctctggc 300
DB 371 ttccacaggctgaagattccaagtatataaaaaaggagactcagggtatccaccgctctggc 430
QY 301 ttgcattcttgcagtttgaggaaactttggagccctcatcatcactgcactccaagc 360
DB
DB 431 ttgcatctttgacgttttgagggaacttttgggagccctcatcatcactgcactccaagc 490
QY 361 aagtattttaaacttcaactactactagagaccacaaaaatggtagaagccttactgtttgggca 420
DB 491 aagtattttaaacttcaactactactagagaccacaaaaatggtagaagccttactgtttgggca 550
QY 421 tctactatattgacacgttttgacattactaaaattgtgtgatgttcagccaatgcag 480
DB 551 tctactatattgacacgttttgacattactaaaattgtgtgatgttcagccaatgcag 610
QY 481 tatttggacctgacttgcagctcttgggcaagcagaagtgagcgagcatcttctt 540
DB 611 tatttggacctgacttgcagctcttgggcaagcagaagtgagcgagcatcttctt 670
QY 541 cttaaggtatgggtggcaccagagaccatttccattcttggagagcttctaatacaagac 600
DB 671 cttaaggtatgggtggcaccagagaccatttccattcttggagagcttctaatacaagac 730
QY 601 ctgttcttgaaggtgatttaagtccatccatcggtcacaaaaatctgcacaaatgcagacatt 660
DB 731 ctgttcttgaaggtgatttaagtccatccatcggtcacaaaaatctgcacaaatgcagacatt 790
QY 661 ttagtctacgataaccattgtctatgtggcgaagatctctgaaggttggagcttcttctaga 720
DB 791 ttagtctacgataaccattgtctatgtggcgaagatctctgaaggttggagcttcttctaga 850
QY 721 atctatagccttcatcaccaacttcaatcaatgaattcagagaaatcagacaattgttaagt 780
DB 851 atctatagccttcatcaccaacttcaatcaatgaattcagagaaatcagacaattgttaagt 910
QY 781 tttagagtttcatcttcacatggaggtaccagttacggttcggtgggaatcagggttcttgcagaa 840
DB 911 tttagagtttcatcttcacatggaggtaccagttacggttcggtgggaatcagggttcttgcagaa 970
QY 841 agtaactctgatgtggatcaactgaaaaaggatttgaatctctgcaaatctgacagccaat 900
DB 971 agtaactctgatgtggatcaactgaaaaaggatttgaatctctgcaaatctgacagccaat 1030
QY 901 cagcattcagatgttatctgtcacaatcagaacctgcagcagcttccagcttgcagatca 960
DB 1031 cagcattcagatgttatctgtcacaatcagaacctgcagcagcttccagcttgcagatca 1090
QY 961 gtatcatatcagaggtagaagaagtgtcaacagctatctgtcacataacttaccagatcat 1020
DB 1091 gtatcatatcagaggtagaagaagtgtcaacagctatctgtcacataacttaccagatcat 1150
QY 1021 cagttattggagagaccaccactatgtgccattttgaaacaaaaagctctccaacaatc 1080
DB 1151 cagttattggagagaccaccactatgtgccattttgaaacaaaaagctctccaacaatc 1210
QY 1081 cgcaccagagcaaaattgaggttcataaagcccgagaagactatttcagctctgttaaaactt 1140
DB 1211 cgcaccagagcaaaattgaggttcataaagcccgagaagactatttcagctctgttaaaactt 1270
QY 1141 cattgccctaaatgtcatttggctgaagaagtccacatgagggcgatttgagataaatt 1200
DB 1271 cattgccctaaatgtcatttggctgaagaagtccacatgagggcgatttgagataaatt 1330
QY 1201 ttctcaggtggtgcactaaaacccagttgtcaagttacaaaaatcacatcattatgat 1260
DB 1331 ttctcaggtggtgcactaaaacccaggtgtcaagttacaaaaatcacatcattatgat 1390
QY 1261 tcaaaaatctgcaccactaaaaatcaaaaaggacgaaaaagtagcaggttcatatttggaaa 1320
DB 1391 tcaaaaatctgcaccactaaaaatcaaaaaggacgaaaaagtagcaggttcatatttggaaa 1450
QY 1321 aataatggttatctcccgctttcaaatgaatgtctacttttggatagaaggaggtacactc 1380
DB 1451 aataatggttatctcccgctttcaaatgaatgtctacttttggatagaaggaggtacactc 1510
QY 1381 agtgaatttgcacaaactctgcacaaagtttaataagtgtaatttctgtgagatctggccac 1440
DB 1511 agtgaatttgcacaaactctgcacaaagtttaataagtgtaatttctgtgagatctggccac 1570

|||||
Db 971 agtaactctgctggtgacactgaaaggattgagatctgcaaatgtgacagccaat 1030
QY 901 cagcattcagatgttatctgtcaatcagaacctgacgacgtttccaaagtctggtatca 960
|||||
Db 1031 cagcattcagatgttatctgtcaatcagaacctgacgacgtttccaaagtctggtatca 1090
QY 961 gtatcatatcagagtgagaagatgtcaacagctctatctgtacataacttacagatcat 1020
|||||
Db 1091 gtatcatatcagagtgagaagatgtcaacagctctatctgtacataacttacagatcat 1150
QY 1021 cagtatgtgagagacacacctatgtgccattttgaaacaaaaagctctccaacaatc 1080
|||||
Db 1151 cagtatgtgagagacacacctatgtgccattttgaaacaaaaagctctccaacaatc 1210
QY 1081 cgcattccagagcaaaattgaggtcatataagccacagaagactatttcagctgttaaaatt 1140
|||||
Db 1211 cgcattccagagcaaaattgaggtcatataagccacagaagactatttcagctgttaaaatt 1270
QY 1141 cattgccctaaatgtcatttctgctcaagaagtccacatgagggcgatcttgatataatt 1200
|||||
Db 1271 cattgccctaaatgtcatttctgctcaagaagtccacatgagggcgatcttgatataatt 1330
QY 1201 ttctcaggatgttgcaactaaaacccagttgtcaagttacaaaatcacatcattatgat 1260
|||||
Db 1331 ttctcaggatgttgcaactaaaacccagatgtcaagttacaaaatcacatcattatgat 1390
QY 1261 tcaaaaaatctgacacactaaaatacaaaaaggacgaaagtacaggttcattttgtgaaa 1320
|||||
Db 1391 tcaaaaaatctgacacactaaaatacaaaaaggacgaaagtacaggttcattttgtgaaa 1450
QY 1321 aataatgttatctcccgctttcaaatgaatgtctacttttgatagaaggaggtacactc 1380
|||||
Db 1451 aataatgttatctcccgctttcaaatgaatgtctacttttgatagaaggaggtacactc 1510
QY 1361 agtgaatttgcaactctgcaactcgaaagttaataatgtaattccctgtgagatctggccac 1440
|||||
Db 1511 agtgaatttgcaactctgcaactcgaaagttaataatgtaattccctgtgagatctggccac 1570
QY 1441 gaagacctggaacttttgacctttcagcaccattctctatacagaaggacaatacatcac 1500
|||||
Db 1571 gaagacctggaacttttgacctttcagcaccattctctatacagaaggacaatacatcac 1630
QY 1501 tatgatgtaaacagtgcttagttgagatccatcacaaaatctaaaatccctgtgattgat 1560
|||||
Db 1631 tatgatgtaaacagtgcttagttgagatccatcacaaaatctaaaatccctgtgattgat 1690
QY 1561 aaaaatcgtgattcccttctctgtgcaagaagcactgggtattgtaccctcccaatat 1620
|||||
Db 1691 aaaaatcgtgattcccttctctgtgcaagaagcactgggtattgtaccctcccaatat 1750
QY 1621 gtgtttgttatgacctttacacttgatgatggaacagagtgactagaagcctatctcatg 1680
|||||
Db 1751 gtgtttgttatgacctttacacttgatgatggaacagagtgactagaagcctatctcatg 1810
QY 1681 gattctgacaaaattctccagatctccagatccagatcagaagttctgagatgagaccttcag 1740
|||||
Db 1811 gattctgacaaaattctccagatctccagatccagatcagaagttctgagatgagaccttcag 1870
QY 1741 aaaaagtggtgatgatgatgatgtgtttgtctccaggaataaaaattgagtcacat 1800
|||||
Db 1871 aaaaagtggtgatgatgatgatgtgtttgtctccaggaataaaaattgagtcacat 1930
QY 1801 ccggtgtgtggaatgtctcatcgaatcacaatgtacaaatgtgcaacagagatacaaat 1860
|||||
Db 1931 ccggtgtgtggaatgtctcatcgaatcacaatgtgcaacagagatacaaat 1990
QY 1861 tgcattcagattttgacacacagtttcagagaagatgtaattctaa 1905
|||||
Db 1991 tgcattcagattttgacacacagtttcagagaagatgtaattctaa 2035

RESULT 3

AAH17603
ID AAH17603 standard; cDNA; 2383 BP.
XX
AC AAH17603;
XX
DT 26-JUN-2001 (first entry)
XX
XX Human cDNA sequence SEQ ID NO:17112.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX
XX 27-AUG-1999; 99JP-0300253.
XX
XX 11-JAN-2000; 2000JP-0118776.
XX
XX 02-MAY-2000; 2000JP-0183767.
XX
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX
XX full-length cDNAs defined in the specification, and for the detection
XX
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX
XX full-length cDNAs -
XX
XX Claim 8; SEQ ID 17112; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesising 5602
XX
XX full-length cDNAs defined in the specification. Where a primer set
XX
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX
XX to the complementary strand of a polynucleotide which comprises one of
XX
XX the 5602 nucleotide sequences defined in the specification, where the
XX
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX
XX of an oligonucleotide comprising a sequence complementary to the
XX
XX complementary strand of a polynucleotide which comprises a 5'-end
XX
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX
XX polynucleotide which comprises a 3'-end sequence, where the
XX
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX
XX the specification. The primer sets can be used in antisense therapy and
XX
XX in gene therapy. The primers are useful for synthesising polynucleotides,
XX
XX particularly full-length cDNAs. The primers are also useful for the
XX
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX
XX AAH13633 to AAH18742 represent human cDNA sequences; AA892446 to
XX
XX AA895893 represent human amino acid sequences; and AAH13629 to AAH13632
XX
XX represent oligonucleotides, all of which are used in the exemplification
XX
XX of the present invention.
XX
XX Sequence 2383 BP; 750 A; 456 C; 459 G; 718 T; 0 other;

Query Match 99.8%; Score 1901.8; DB 22; Length 2383;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1903; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 atgtcttgggtccagcaacaaattatatatatatacccccgaatcaacttaagggtggt 60
|||||
Db 439 atgtcttgggtccagcaacaaattatatatatatacccccgaatcaacttaagggtggt 498

PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 XX full-length cDNAs -
 PS Claim 8: SEQ ID 11187; 2537pp + CD ROM; English.
 XX The present invention describes primer sets for synthesising 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 2631 BP; 826 A; 473 C; 491 G; 841 T; 0 other;

Query Match 100.0%; Score 1905; DB 22; Length 2631;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1905; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 atgtcttgggtccagcaacaattatatatacacacccctgaatcaacttaagggtggt 60
 Db 24 atgtcttgggtccagcaacaattatatatacacacccctgaatcaacttaagggtggt 83
 Qy 61 acaattgcaatgctatggtgtgtgaagttctttaagcccccatatcttaagcaaaaga 120
 Db 84 acaattgcaatgctatggtgtgtgaagttctttaagcccccatatcttaagcaaaaga 143
 Qy 121 actgattattgctcagttgtaactattgtgaccagacacaaatgtataaactaacttgctcg 180
 Db 144 actgattattgctcagttgtaactattgtgaccagacacaaatgtataaactaacttgctcg 203
 Qy 181 ctcttagtggaaactatgaagccctcccaataattataaaatggagatatgttctgcg 240
 Db 204 ctcttagtggaaactatgaagccctcccaataattataaaatggagatatgttctgcg 263
 Qy 241 ttccacagctgaagattcaagtatataaaaggagactcaggttatccagctctggc 300
 Db 264 ttccacagctgaagattcaagtatataaaaggagactcaggttatccagctctggc 323
 Qy 301 ttfgcatctttgacgtttgagggaactttggagccctcattacacacctcgcaacttcaagc 360
 Db 324 ttfgcatctttgacgtttgagggaactttggagccctcattacacacctcgcaacttcaagc 383
 Qy 361 aagattttaacttcaactactgaggaacacaaaatggttagaagccttaagtttgggca 420
 Db 384 aagattttaacttcaactactgaggaacacaaaatggttagaagccttaagtttgggca 443
 Qy 421 tctactcatatgcacgtcttgacattactataaattgtgtgatttcacgaccaatgcag 480
 Db 444 tctactcatatgcacgtcttgacattactataaattgtgtgatttcacgaccaatgcag 503
 Qy 481 tattttgacctgacttgcagctcttgggcaaaagcagaagtggaagcagcatcttctt 540
 Db 504 tattttgacctgacttgcagctcttgggcaaaagcagaagtggaagcagcatcttctt 563
 Qy 541 ctaaagggtatggatggcaccaggacaccatttccattcttggagagtcttaatacaagac 600

Db 564 ctaaagggtatggatggcaccaggacaccatttccattcttgggagcttctaatacaagac 623
 Qy 601 ctgtcttgaaggtgatttaagtacatccatcggtacacaaatctgacaatagacatt 660
 Db 624 ctgtcttgaaggtgatttaagtacatccatcggtacacaaatctgacaatagacatt 683
 Qy 661 ttaagtacgataaaccatgttcatgtgcaagatctctgaagttggaagcttctttaga 720
 Db 684 ttaagtacgataaaccatgttcatgtgcaagatctctgaagttggaagcttctttaga 743
 Qy 721 atctatagcttctacacaaacttcaatcaatgaattcagagaaatcagacaattgttaagt 780
 Db 744 atctatagcttctacacaaacttcaatcaatgaattcagagaaatcagacaattgttaagt 803
 Qy 781 ttaagtttcatcttcatggaggtaccagttacaggttcgggggaatcagggtcttgcagaa 840
 Db 804 ttaagtttcatcttcatggaggtaccagttacaggttcgggggaatcagggtcttgcagaa 863
 Qy 841 agtaactctgattggtacaaactgaaaagatttagaacttgcacaaatttgacagccaat 900
 Db 864 agtaactctgattggtacaaactgaaaagatttagaacttgcacaaatttgacagccaat 923
 Qy 901 cagcattcagatgttatctgtcattcagaaacctgacgacagcttccaagctctggatca 960
 Db 924 cagcattcagatgttatctgtcattcagaaacctgacgacagcttccaagctctggatca 983
 Qy 961 gtaaatattacgaggtagaagaatgtcaacagctatctgtacaatacttacaagatcat 1020
 Db 984 gtaaatattacgaggtagaagaatgtcaacagctatctgtacaatacttacaagatcat 1043
 Qy 1021 cagtatttggagagacaccactatgtgccattttgaaacaaaagctctccaacaatcac 1080
 Db 1044 cagtatttggagagacaccactatgtgccattttgaaacaaaagctctccaacaatcac 1103
 Qy 1081 cgcctcagcagcaaatggaggtcatataagcccagagagactatttcagctctgttaaacctt 1140
 Db 1104 cgcctcagcagcaaatggaggtcatataagcccagagagactatttcagctctgttaaacctt 1163
 Qy 1141 cattgcctaaatgtcatttctgctcaagaagtccacatgagggcgatttggatataatt 1200
 Db 1164 cattgcctaaatgtcatttctgctcaagaagtccacatgagggcgatttggatataatt 1223
 Qy 1201 ttccagatgggtgcaactaaaccccgattgtcaagtgtacttttgaagaaggaggtacactc 1260
 Db 1224 ttccagatgggtgcaactaaaccccgattgtcaagtgtacttttgaagaaggaggtacactc 1283
 Qy 1261 tcaaaaaatctggaccactaaaaatcaaaaaggacgaaaagtagcagttcattttgtgaaa 1320
 Db 1284 tcaaaaaatctggaccactaaaaatcaaaaaggacgaaaagtagcagttcattttgtgaaa 1343
 Qy 1321 aataatggtattctcccgctttcacaatgaatgtctacttttgaagaaggaggtacactc 1380
 Db 1344 aataatggtattctcccgctttcacaatgaatgtctacttttgaagaaggaggtacactc 1403
 Qy 1381 agtgaatttgcacaaactctgaacaaagttaataagttaattctctgtgagatcttggccac 1440
 Db 1404 agtgaatttgcacaaactctgaacaaagttaataagttaattctctgtgagatcttggccac 1463
 Qy 1441 gaagacctgggaacttttgaccttccagcaccatttcttatacagaagacaatacatcac 1500
 Db 1464 gaagacctgggaacttttgaccttccagcaccatttcttatacagaagacaatacatcac 1523
 Qy 1501 tatggtataaacaggttcttagtttgagatcccatcaaaaatctaaatctccctggtgat 1560
 Db 1524 tatggtataaacaggttcttagtttgagatcccatcaaaaatctaaatctccctggtgat 1583
 Qy 1561 aaaaactcgtgattctcttcttctgtggcagaagcactgggtattgtaccctccaatcat 1620
 Db 1584 aaaaactcgtgattctcttcttctgtggcagaagcactgggtattgtaccctccaatcat 1643
 Qy 1621 gtgtgttattgacatttcaacttgcattgatgtgaacagagactagaagccttatctcatg 1680

PT peripheral neuropathy -

XX Example 2; Page 90-93; 126pp; English.

XX The present sequence represents module 8 of the FK-520 polyketide
CC synthase (PKS) gene cluster of strain MA6548. FK-506 is a potent
CC immunosuppressant, and acts through initial formation of an
CC intermediate complex with protein immunophilins known as FK-506
CC binding proteins. The nucleic acids are used for producing polyketide
CC compounds. The polyketide compounds can be used as immunosuppressants to
CC prevent or treat transplant rejection, graft-versus-host disease or
CC uveitis. They can also be used for treating e.g. alopecia universalis,
CC autoimmune chronic active hepatitis, inflammatory bowel disease,
CC multiple sclerosis, primary biliary cirrhosis, or scleroderma. They
CC also have neurotrophic activity and can be used to promote neurite
CC outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
CC and in intact animals, they promote regrowth of damaged facial and
CC sciatic nerves, and repair lesioned serotonin and dopamine neurons in
CC the brain. They can also be used for treating e.g. Parkinson's disease,
CC Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
CC peripheral neuropathies. They can also be used in agricultural and
CC veterinary applications.

XX Sequence 1574 AA:

Query Match 3.2%; Score 105.5; DB 21; Length 1574;
Best Local Similarity 21.9%; Pred. No. 2.2;
Matches 92; Conservative 44; Mismatches 160; Indels 125; Gaps 21;

QY 167 QLLGKAEVDC-----SFLKVDGTRTPPSHRVLI-----QDLVLEGDLSHIH 211
DB 1130 qvfveavdspdgfvahpdlldavfsavgdgsrqp-tgwrldavhasdatviracit--- 1185
QY 212 RLQNLTIIDILVDN---HVHVARSLKVGSLRIYSLHKLQSMNSNQTMLSLEFHLHGG 268
DB 1186 rrdsgvvelaafdgagmpvltasvltg-----evasaggsdesdglrllew----- 1232
QY 269 TSYGRGIRVLPSNSVDQLKRDLESANLTANOHSVDVICOSEPDSDFFPSSGVSLEYEVER 328
DB 1233 -----lpvaeahydgadelpegytllitathpd-----dpdptnphmtpttrhtqt 1278
QY 329 CQOLSA-----TILTDHOYLERT---PLCAI--LKOKAPQOYRIRAKL-----RSYKPRRL 374
DB 1279 trvltalqhlitnhtlvttdppgaavtgltrtaqnehpgrihliethhphptlpl 1338
QY 375 FQSVKLHCPKCHLLQEPHGGDLIIIFQDGATKTPVVKLQNTSLYDSKIWTTKNQKGRKV 434
DB 1339 tqittlhqphlrltnntlhtphl-----tpitthhntt-----tptntp---- 1378
QY 435 AVHFVKMNGILPLSNECLLIEGGLTSEICKLSKNFNSVIPVRSGHEDLELDLS-APFL 493
DB 1379 -----plnphallitgg-----sgtliagilarhlnhphtyllsrtppppt 1419
QY 494 IOGTIHHYCKQCSLSRLNSLNSVDKTSWIPSSVAEALGIV--PLQYVFWMTFLDDG 551
DB 1420 tpgt--hipc-----dltd-----ptqitqalthipqplgtgftaataidda 1459
QY 552 T 552
DB 1460 t 1460

Search completed: April 9, 2002, 17:08:10
Job time: 589 sec

PR	16-JUL-1999;	99US-0144086;
PR	19-JUL-1999;	99US-0144325;
PR	19-JUL-1999;	99US-0144331;
PR	19-JUL-1999;	99US-0144332;
PR	19-JUL-1999;	99US-0144333;
PR	19-JUL-1999;	99US-0144334;
PR	19-JUL-1999;	99US-0144335;
PR	20-JUL-1999;	99US-0144352;
PR	20-JUL-1999;	99US-0144632;
PR	20-JUL-1999;	99US-0144684;
PR	21-JUL-1999;	99US-0144814;
PR	21-JUL-1999;	99US-0145086;
PR	21-JUL-1999;	99US-0145088;
PR	22-JUL-1999;	99US-0145087;
PR	22-JUL-1999;	99US-0145089;
PR	22-JUL-1999;	99US-0145192;
PR	23-JUL-1999;	99US-0145145;
PR	23-JUL-1999;	99US-0145210;
PR	23-JUL-1999;	99US-0145224;
PR	26-JUL-1999;	99US-0145276;
PR	27-JUL-1999;	99US-0145913;
PR	27-JUL-1999;	99US-0145918;
PR	27-JUL-1999;	99US-0145919;
PR	28-JUL-1999;	99US-0145951;
PR	02-AUG-1999;	99US-0146386;
PR	02-AUG-1999;	99US-0146388;
PR	02-AUG-1999;	99US-0146389;
PR	03-AUG-1999;	99US-0147038;
PR	04-AUG-1999;	99US-0147204;
PR	04-AUG-1999;	99US-0147302;
PR	05-AUG-1999;	99US-0147192;
PR	05-AUG-1999;	99US-0147260;
PR	06-AUG-1999;	99US-0147303;
PR	06-AUG-1999;	99US-0147416;
PR	09-AUG-1999;	99US-0147493;
PR	09-AUG-1999;	99US-0147935;
PR	10-AUG-1999;	99US-0148171;
PR	11-AUG-1999;	99US-0148319;
PR	12-AUG-1999;	99US-0148341;
PR	13-AUG-1999;	99US-0148565;
PR	13-AUG-1999;	99US-0148684;
PR	16-AUG-1999;	99US-0149368;
PR	16-AUG-1999;	99US-0149175;
PR	18-AUG-1999;	99US-0149426;
PR	20-AUG-1999;	99US-0149722;
PR	20-AUG-1999;	99US-0149723;
PR	20-AUG-1999;	99US-0149929;
PR	23-AUG-1999;	99US-0149902;
PR	23-AUG-1999;	99US-0149930;
PR	25-AUG-1999;	99US-0150566;
PR	26-AUG-1999;	99US-0150884;
PR	27-AUG-1999;	99US-0151065;
PR	27-AUG-1999;	99US-0151066;
PR	27-AUG-1999;	99US-0151080;
PR	30-AUG-1999;	99US-0151438;
PR	31-AUG-1999;	99US-0151303;
PR	01-SEP-1999;	99US-0151930;
PR	07-SEP-1999;	99US-0152363;
PR	10-SEP-1999;	99US-0153070;
PR	13-SEP-1999;	99US-0153758;
PR	15-SEP-1999;	99US-0154018;
PR	16-SEP-1999;	99US-0154039;
PR	20-SEP-1999;	99US-0154779;
PR	22-SEP-1999;	99US-0155139;
PR	23-SEP-1999;	99US-0155486;
PR	24-SEP-1999;	99US-0155659;
PR	28-SEP-1999;	99US-0156458;
PR	29-SEP-1999;	99US-0156596;
PR	04-OCT-1999;	99US-0157117;
PR	05-OCT-1999;	99US-0157753;
PR	06-OCT-1999;	99US-0157865;
PR	07-OCT-1999;	99US-0158029;

PR	08-OCT-1999;	99US-01582323
PR	12-OCT-1999;	99US-01583669
PR	13-OCT-1999;	99US-01592993
PR	13-OCT-1999;	99US-01592994
PR	13-OCT-1999;	99US-01592995
PR	13-OCT-1999;	99US-01593299
PR	14-OCT-1999;	99US-01593330
PR	14-OCT-1999;	99US-01596337
PR	14-OCT-1999;	99US-01596338
PR	18-OCT-1999;	99US-01599364
PR	21-OCT-1999;	99US-01600741
PR	21-OCT-1999;	99US-01600767
PR	21-OCT-1999;	99US-01600768
PR	21-OCT-1999;	99US-01600770
PR	21-OCT-1999;	99US-01600814
PR	21-OCT-1999;	99US-01600815
PR	22-OCT-1999;	99US-01600980
PR	22-OCT-1999;	99US-01600981
PR	22-OCT-1999;	99US-01600989
PR	25-OCT-1999;	99US-01614004
PR	25-OCT-1999;	99US-01614005
PR	25-OCT-1999;	99US-01613599
PR	26-OCT-1999;	99US-01613660
PR	26-OCT-1999;	99US-01613661
PR	28-OCT-1999;	99US-01619920
PR	28-OCT-1999;	99US-01619992
PR	28-OCT-1999;	99US-01619993
PR	29-OCT-1999;	99US-01621442

Query Match

Query Match	5.2%;	Score 106;	DB 21;	Length 1123;
Best Local Similarity	20.4%;	Pred. No. 1.1;		
Matches 104;	Conservative	65;	Mismatches 169;	Indels 172;
	Gaps	25;		

```

118 TSSYFNFTTEDHKWEAL---RWASTHSPSWTLKLCDVQPMOYFDLT-----165
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
62 tpcnwfgitcdsknvaslnfrsrvsqqiqp-----eigelksqildstnnfsgtip 116
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
166 ----COLGKAEVDGASFLKKVYWDGTRTFPPSRVLIODL-VLEGDLS----HIURLQ- 214
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
117 stlgncktklatldisengfsdkipd-tldaklrlevlylvinfltgelpeslfrlklqv 175
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
215 ----NLT-----DILYVDNHV--HVARSILKVGSRFLRIVSLH-----245
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
176 lyldynnltpipqsgidakelvelsmyanqfsgnipesignsssiqilylhrnkivgsl 235
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
246 ----TKLQSMSENOTMLSLFEH--LHGQTSYVGRCIRVLPPES 281
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
236 pesinllgnltlrvgnnslqqpvrfsqpncknlltldlsynefegvvpalg-----n 289
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
282 NSDVDQL-----KKDLESANLTANOHSDIVICQEPDDSFSPSSGVSVLY 324
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
290 cssidalvivsgnlsqtipsalgnaknltlinisen-----lsgslpa- 333
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
325 EVERCOOLSATILTDHOYLERTPLCAILKQAKPOQYIRAKLSYKPRRLFQSVKLHCPR 384
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
334 eigncsslnllkndnqlvgvipsal-----gklrkleselfen-----373
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
385 CHLQEVPEHGDLDIIFODGATKTPVVKLQNTSLYDSKIWTTRKQGRKVAVHFVNN---442
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
374 ----rfsgeipleiawksqslqllvygnnltegeipventemkk-lklatif--nnsf 423
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
443 -GILPLSNECLLLTEGGTSLPICKLSNKNFVPIVRSGH--EDLELLDLSAPPL---IQGT 497
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
424 ygaippg-----lgnvsleevdfignkltegeipnlchgrkriiolsnllhtgipas 478
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
498 IHHYGCKQCCSLRSI----QNLNSLVDKTS 523
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
479 igh-----cktirrfilrennlsallpefs 503
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 14

Query Match 10.8%; Score 358; DB 22; Length 66;
Best Local Similarity 100.0%; Pred. No. 1.1e-27;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 569 PASEVLMDDDLQKSDYMDNFCPCGKIDAYPWLECFIKSYNVTNGDNOICYQIFDIT 628
|||||
Db 1 pasevlmdddlqsksdymndmfcpggikidaypwlecfiksyntngdnqicyqifdt 60

QY 629 VAEDVI 634
|||||
Db 61 vaedvi 66

RESULT 7
AAM34264
ID AAM34264 standard; Protein; 30 AA.
AC AAM34264;
XX
XX 17-OCT-2001 (first entry)
XX
DE Peptide #8301 encoded by probe for measuring placental gene expression.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200157272-A2.
PN
PD
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207450.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
XX Claim 27; SEQ ID No 34533; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP;
CC see AA13115-AA157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
XX
XX Sequence 30 AA;
SQ

Query Match 4.4%; Score 147; DB 22; Length 30;
Best Local Similarity 96.7%; Pred. No. 1.9e-07;
Matches 29; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 389 QEVPHGGDLIIIFODCATKTPVVKLQNTSL 418
|||||
Db 1 qevphgdlilifodgatktpvkvkqlntsl 30

RESULT 8

AAG58505
ID AAG58505 standard; Protein; 467 AA.
XX
AC AAG58505;
XX
XX 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 75533.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
OS
XX EPI033405-A2.
PN
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 23-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.

```
SQ Sequence 634 AA:
Query Match 99.8%; Score 3317; DB 22; Length 634;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 633; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSLVPATNYIYTPNLKGGTIVNVYGVVKKFPPYLSKGTIDYCSVVTIVDQTNVLTCL 60
DB 1 mslvpatnyiytpnlqkggtivnvvgvkvffkppylskgtidycsvvtivdqtvnkltcl 60
QY 61 LFSGNYEALPIIYKNGDIVRFHRLKIQYKKETQGTSSGFSASLTGEGTLGAPIIPRTSS 120
DB 61 lfsngyealpilykngdivrfhrklyqykktqgtssgsfsltgegtlgapiiprtss 120
QY 121 KYNFETEDHKMVEALRWASTHMSPTLLKLCVOPMQYFBLTCLGLGKAEVDCASFL 180
DB 121 kyinfetdtkmvealrwasthmspswtllklcdvqpmqyfdltcqlglkaevdgasfl 180
QY 181 LKVDGTRTPFPSSRWVLIQDLVLEGLSHIHLQNLITDILVYDNHVVHVARSLKVGSLR 240
DB 181 lkwdgtrtpfssrwlqdlvleghslhrlqnlitdilvydnhvvhvarsikvgsflr 240
QY 241 IYSLHTKLQSMNSNOTMLSLEFHLHGGSYGRGIRVLPESNSDVQDLKKDLESANLTAN 300
DB 241 iyslhtklqsmnsenqtmlslefhlggtsgygrgirvlpesnsdvdqldkldlesanltan 300
QY 301 QHSDVICQSEPDSPSSGVSLSYEVERCOOLSATILTDHOYLERTPLCALIKKAPQY 360
DB 301 qhsdvicqsepdspssgsvslyevercqolsatiltldhoylertplcalikqkapy 360
QY 361 RIRAKRSYPRRLFSQVSKLHCPKCHLLQEPVHPHGDIIIFQDGATKTPVVKLQNTSLYD 420
DB 361 rirakrsyprrlfsgvsklhcpkchllqevhphegdliifdgatktpvvklqntslyd 420
QY 421 SKIWTNKGKRVAVHFVNNGILPLSNCELLIEGGTILSEICKLSKNFNSVIPVRSKH 480
DB 421 skiwtknqgrkvavhfvnngilplsnecellieggtilseicklsknfnsvipvrsgh 480
QY 481 EDLELIDLSAPFLIOGTIHHYCKQCSLSRSLQNLNSLDVKTSMIPSSVAEALGIVPLGY 540
DB 481 edlelidlsapfliogtthhyckqcsllsrslqnlslsdvktswipssvaealgivply 540
QY 541 VFVMTITLDGTVLEAYLWDSKFFQIPASEVLMDDDLQKSVDMIMDMFCPPGIKIDAY 600
DB 541 vfvmtitldgvtvleaylmdskffqipasevlmdddlqksvdmimdmfcppgikiday 600
QY 601 PWLECFIKSVNTGTDNOICYOIFEDTTVAEDVI 634
DB 601 pwlecfiksvntgtdnolcycyofedttvaedvi 634

RESULT 3
AAB93478
ID AAB93478 standard; Protein; 634 AA.
XX
AC AAB93478;
XX
XX 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:12761.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
XX
PD 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
PF
XX 29-JUL-1999; 99JP-0248036.
PR
```

```
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs.
PS Claim 8; SEQ ID 12761; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX Sequence 634 AA:
SQ
Query Match 99.8%; Score 3317; DB 22; Length 634;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 633; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MSLVPATNYIYTPNLKGGTIVNVYGVVKKFPPYLSKGTIDYCSVVTIVDQTNVLTCL 60
DB 1 mslvpatnyiytpnlqkggtivnvvgvkvffkppylskgtidycsvvtivdqtvnkltcl 60
QY 61 LFSGNYEALPIIYKNGDIVRFHRLKIQYKKETQGTSSGFSASLTGEGTLGAPIIPRTSS 120
DB 61 lfsngyealpilykngdivrfhrklyqykktqgtssgsfsltgegtlgapiiprtss 120
QY 121 KYNFETEDHKMVEALRWASTHMSPTLLKLCVOPMQYFBLTCLGLGKAEVDCASFL 180
DB 121 kyinfetdtkmvealrwasthmspswtllklcdvqpmqyfdltcqlglkaevdgasfl 180
QY 181 LKVDGTRTPFPSSRWVLIQDLVLEGLSHIHLQNLITDILVYDNHVVHVARSLKVGSLR 240
DB 181 lkwdgtrtpfssrwlqdlvleghslhrlqnlitdilvydnhvvhvarsikvgsflr 240
QY 241 IYSLHTKLQSMNSNOTMLSLEFHLHGGSYGRGIRVLPESNSDVQDLKKDLESANLTAN 300
DB 241 iyslhtklqsmnsenqtmlslefhlggtsgygrgirvlpesnsdvdqldkldlesanltan 300
QY 301 QHSDVICQSEPDSPSSGVSLSYEVERCOOLSATILTDHOYLERTPLCALIKKAPQY 360
DB 301 qhsdvicqsepdspssgsvslyevercqolsatiltldhoylertplcalikqkapy 360
```

RT orf1, orf2, orf3, orf4 (DNA segment VD1) [parvo-like virus, Yamashita
RT isolate, host-silkworm, Genomic, 6542 nt].
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB033596; BA85362.1; -
KM Nonstructural protein.
SQ SEQUENCE 1115 AA: 128304 MW: 7C0DA2DCFA746428 CRC64:

Query Match 4.2%; Score 115; DB 12; Length 1115;
Best Local Similarity 19.8%; Pred. No. 0.79;
Matches 122; Conservative 93; Mismatches 186; Indels 216; Gaps 35;

OY 5 PATNYITP-----LNO-----LKGITV-----NYGVVFFKPPYLSKGTDCSV 46
DB 363 PAEMFIFKSYLANLEDEVFNDEGEIIFNDIEMFLTDLSIIRKFOI-----NDYGKE 416
OY 47 VTIVDQINVLKTC-----LFGSGVEALPIIYNGDIVRFRHLK-- 85
DB 417 ANLIDGGERKCLACKNNKKGYVARHNGLIFGYCRSCFLAHNNTPIVYHFKGY 476
OY 86 -----IOVKKETOGITSSG-----FASLTFEGTLGAPIIPRTSSKY 122
DB 477 DHIHLDELKDKSKHNTCRGSKINKMDVITHKDLDSFIITFKDI----- 523
OY 123 FNFTEDHKVEALRWASTHMSPTLLKLCV-----QPOYFDLTCOLLGKAEV 174
DB 524 FNF-----LPESLASLANKLTLTKYTPDKFKDAFNSGKGFPEWED--DFNKLEIEVP 576
OY 175 -----DGASFILK-----VW-DGTRTFPSMRVLIOD-----LVLE----- 204
DB 577 ODPADMSRLTNKKGTETIIKKANOIWDNNMOIFHEVLYLNLDELWLLLEVEAFRTD 636
OY 205 -----GDLSHIRLQNLTDI--LVYDNVH--VARSLSGSEFLRIYSLH 245
DB 637 TVNEDKIDPVYFDGAPGLTFELARMYESSLDMHYIPDKNYLDVSRIRVG-----VIOVV 692
OY 246 TKLOSMSSENOTMLSEFLHSGTSGRGIRVLPESNSDVQKKDESAULTANQMSDV 305
DB 693 TKYANIEDVDETIYLLD--VNTMYSTCMKOKLANKYLGTDITLTDNYS-----DDN 742
OY 306 ICO--SEPDPSFGSSGSVLYEVERCOOLSATI--LTDHOYLERTPLCALIKOKAPOYR 361
DB 743 FCYFIKGDFTSP-----EYLHDLPAHLSMPLMHQV--NNKLCITFLDK--KDL 788
OY 362 IRAWLSRTKPRRLFOVKYKLCRCHLQOEVPH-----GDLDIIFDDGATK--TPD-VK 412
DB 789 IHSKVFYR-----YLSGLVCDDKIHYVYKFKOEYIKDYVEINIOKRNSSTIDPGIKDYK 843
OY 413 LQNTSLVDS-----KIMTKRNOG-----RKVAVHFVYNNGLPLSNCL--LL 454
DB 844 LKNNALGCKTCENFKRKISFVTNVNSGDREKNCMSKAKSHITLGNCL--LYECVIRYL 901
OY 455 IE-----GGTISEITCKL 466
DB 902 LDKPIQIGFTILELAKL 918

RESULT 12
O9C2K8 PRELIMINARY; PRT; 759 AA.
AC O9C2K8;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE CONSERVED HYPOTHEITICAL PROTEIN.
GN 3410.40.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_Taxid-5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,

RA Nyakatura G., Meves H.W., Mannhaupt G.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL513442; CAC28643.1; -
SQ SEQUENCE 759 AA: 85218 MW: DC73A5574FF4F73 CRC64:

Query Match 4.1%; Score 111; DB 3; Length 759;
Best Local Similarity 20.7%; Pred. No. 0.99;
Matches 63; Conservative 50; Mismatches 125; Indels 66; Gaps 12;

OY 3 LVPAITNYIRPLNQ-----KGGTIYNYGVYKFFKPPYLSKGTDCSVYTI-----VD 51
DB 18 LFPST--LTPRLALIDDGDAQGSHVNYIGLKDCAVAVAHGSDMKCTLTISLSIE 74
OY 52 QINVKLTCLFSGVEALPIIYKNGDIVRFRHLKIQYKKETOGITSSGFASLTFTGLG 111
DB 75 DESAGVELVIF--RPEARMEVGAGDVLVLSAKVQRFKNPSLITSKITTCVYKKAFT 132
OY 112 APII-----PRTSKYFNFTEDHKVEAL-----RYMASTHM 144
DB 133 IPVYPASAOYALLPPKGGESHKLKEEHQYVSILYVNDKYDVPDEAEYQQRVYKSLNV 192
OY 145 SPSTVTLKLDVOPMOYFDLTCOL-----LQKAEV-----DGASFILKWDGTRTP 190
DB 193 KDRFSLLK--DIVDGNRYDLIGQVAKPDYDEMOKRILYLSQDTENDLFHHYMEGVRDL 250
OY 191 FPSNR--VLIODLVLEGLSH--IHRLONLTDILVYDNHVAR--SLKVSFLRIYS 243
DB 251 ASAARPADAYVEDNNPMAOOHPVYGVGRKTIIDISCYDAHADFIIRAGVSAGMWLSLRN 310
OY 244 LHTK 247
DB 311 VOYK 314

RESULT 13
O9XE80 PRELIMINARY; PRT; 483 AA.
AC O9XE80;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE SERINE-TYPE CARBOXYPEPTIDASE.
OS Sorghum bicolor (Sorghum) (Sorghum vulgare).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Sorghum.
OX NCBI_Taxid-4558;
RN [1]
RP SEQUENCE FROM N.A.
RA Liaca V., Lou A., Messing J.W.;
RT "Microsynteny analysis of 22-kda zein cluster in maize and sorghum."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF061282; AAD22150.1; -
DR HSSP: P08819; 1WHT.
DR InterPro: IPR000379; Est_1ip_thioest_actsite.
DR InterPro: IPR001563; Serine_carptsept.
DR Pfam: PF00450; serine_carptsept. 1.
DR PRINTS: PR00724; CRBOXYPTASEC.
DR PROSITE: PS00131; CARBOXYPEPT_SER; UNKNOWN_1.
KM Carboxypeptidase.
SQ SEQUENCE 483 AA: 53413 MW: DB91DCGEF772A0D9 CRC64:

Query Match 4.0%; Score 110.5; DB 10; Length 483;
Best Local Similarity 20.6%; Pred. No. 0.57;
Matches 91; Conservative 69; Mismatches 134; Indels 147; Gaps 28;
OY 95 GITSSGFASLTFTGLCAPIIIPRTSKYFNFTEDHKVEALRWASTHMSPTLLKLC 154


```
Db 655 LEEVHHSLGCCSKYIGLYLNDOCKSLKRP-CVNVSELEYLGLRSCDSLEKLPETIYGRMKP 713
OY 365 KLRSTK-----PRLLFOSVKLHCPK-----CHL 387
Db 714 EYQIHMOGSGIRELPSSIFQ-YKTHVTKLLMNMKNLVALPSSICRLKSLVSLVSGCSK 772
OY 388 LOEVPHE-GDLDDIFODGATKTPDVKIONTSLYDSKIWTIRKNOGRKVAHF-----438
Db 773 LESLPEEIGDLDNLRVFDASDTLILRPPSSILRLNKL-IILMFRGFKDGVHFEFPVAEG 831
OY 439 YKNNGILPLSNCELLIEGTLSEICLSNKFNSYIPYRSGHEDELDLS 489
Db 832 LHSLEYLNL--YCNLDIDGGLPEEIGSLSS-----LKKLDELS 865
```

Search completed: April 9, 2002, 17:08:56
Job time: 275 sec

STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: LVZNMOT01
 CLONE: 348429
 US-09-274-570-3

Query Match 3.3%, Score 91; DB 3; Length 425;
 Best Local Similarity 19.5%; Pred. No. 0.96;
 Matches 92; Conservative 81; Mismatches 175; Indels 124; Gaps 22;

OY 23 VAVYGVAFKFP--YLSKGD--YCSVYIVDOINVKLTCL---LFGNNEALPIIKN 75
 DB 16 LSLSGVGFDSLPDOLVNRKSTSGCFNLCVGETGICKSTLMDLFTNKFESDEPATNE 75
 OY 76 GDVHFHRLKIOVYKKEGTGISTSGEASLTFEGTLG-----APIIPRTSKYEN 124
 DB 76 PGV---RLKARSTYLOESNR---LKLTIYIVGFGDQINKDSYKFIYVIOAEFA 127
 OY 125 FTEDHKVYAL---RYMASTH--MSPSWTLKLCDOVQPYFDLTCOLGKAEVDGA 177
 DB 128 YLQEFELKIRSLFNHYDRIRHACLYFIAPTGHSLKSLDLVTMKLDS----- 174
 OY 178 SFLKVMGRTFRPSWVLLIODLVLEGD---SHIHLQNTLIDILYDNHVAARSLK 234
 DB 175 -----RVNIIPILAKADTIAKNELHFKSKIMSELVNG----- 208
 OY 235 VGSFLRIYSLTKLOSMNSENOTM--LSLEFHLGCT-----SYGRG--IRVLP 279
 DB 209 ---VQITQFPTDETVAINATMSVHLFPAVYGSTEEKIGNKAKARQYMWGVQVEN 264
 OY 280 ESNVDOLKDLKDESANL-----TANQSDVI--COSEP---DPSFGSGVSL---YE 325
 DB 265 ENHCDFVRLREMLIVNNEDELREQTHRTHYELRYCKLEEMGFKDTDDSKPFSLOETYE 324
 OY 326 VERCOOLSATILIDHOYLETPLCALILKOKAPO---OYIRAKLRSKPRFLFQSVLH 381
 DB 325 AKRNFLE--LQKKEEPMOFYWRKVEKEALKEAELHEKDLKRTIHOEKKVE 382
 OY 382 CPKCHLOEVPHEGDLIIFODGATKTPDVKLONTSLYDSKIWTITKNOGRK 433
 DB 383 DKRKELEVNN-----FQ---KKRAAOULLSOAQSAGQOKKXKDKK 424

RESULT 13

US-08-323-1708-2
 Sequence 2, Application US/083231708
 Patent No. 5733772

GENERAL INFORMATION:
 APPLICANT: Williamson, Kim C.
 TITLE OF INVENTION: Cloning and Expression of Plasmodium
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESS: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, 8th Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/323,1708
 FILING DATE: 13-OCT-1994
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/010,409

FILING DATE: 29-JAN-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Quine, Jonathan A.
 REGISTRATION NUMBER: P-41,261
 REFERENCE/DOCKET NUMBER: 015280-113100US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3135 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-323-1708-2

Query Match 3.3%, Score 91; DB 1; Length 3135;
 Best Local Similarity 18.3%; Pred. No. 29;
 Matches 105; Conservative 89; Mismatches 185; Indels 194; Gaps 28;

OY 20 GIIVN---VGV---VFEKPPYLSKGTDY-----CSVYTIYDQTN-VKLTGLESGN 65
 DB 825 GNLVNSVYVKNEMAKFNOYVHIPISTYKDTLNLFCIILKEESNLSTSLYVSI 884
 OY 66 YEALPIIKNGDIYRFH-----RLKIOYRK-----ETOGITSSG- 100
 DB 885 NEEL-----NFSLFDYFESFVPIKKTIOVAKRNNKEDHYTCDFTDKDKTVPSTANG 939
 OY 101 -----FASLTFEGTL-----GAPLIIPRT-----SSKYENF 125
 DB 940 KLFICRKLKEEDFTTLCCGNKTOYPIELIPPLKDKKVLKLDLDIOYOMFSKFFK 999
 OY 126 TTEDHKVYEA---LRMASTH-----SPSWTLKLCDOVQPYFDLTCOLGKAEVD 175
 DB 1000 NTQNAKYLVLPYLIIPFNHIGKELKNPTK-----NKKDVKYFEGSSVLSPLSSAD 1054
 OY 176 GASFLKAWDGTFRPPSPRWVLLIODLVLEGDLSHIHLQNTLIDILYDNHVAARSLK 235
 DB 1055 SLGKLNLFDLOET---VCLTEKI-----RYLNLSTINELGSDNNT--FSVYFOY 1098
 OY 236 GSFRLRIYSLTKLOSMNSENOTMLSLEFHLGCTSYGRGIRVLPESNSDDOL----- 288
 DB 1099 PPIYDIKE-----PFYFMFGCANNKG-----EGNIGIVELLISKOE 1135
 OY 289 -----KKDESANLTANQH-----SDVICSEPDSPSSGSVSYEVERCO 330
 DB 1136 KIKGCFHESKIDYFNEMISDTHCETLHAYENDIIGFNCLETTHPNEVEVEYEDAEIYL 1195
 OY 331 OLSATILIDHOYLETPLCALILKOKAPOQYRIRAK-----LSYKPRRLFQSYKILKCKH 386
 DB 1196 QPENCFNNVYKGLNSVDITTLKN--AOTYININKKTPFTFLKIPPYNLLEDEVEISC--OCT 1252
 OY 387 LLOEVPHEGDLIIFODGATKTPDVKLONTSLYDSKIWTITKNOGRKVAHVAVNNGILP 446
 DB 1253 IKQYVK---KIKVITTKNDYVLKREVQSESTLDDKIKCKHE-----NFINP 1297
 OY 447 LSNCELLIEGTLSE---IC--KLSKRENSV 473
 DB 1298 RVNK-----TFDENVEYTCNIKIEFENFYI 1322

RESULT 14

US-08-310-912A-108
 Sequence 108, Application US/08310912A
 Patent No. 5981730

GENERAL INFORMATION:
 APPLICANT: Ausubel, Frederick M.
 APPLICANT: Staskawicz, Brian J.
 APPLICANT: Brent, Andrew F.
 APPLICANT: Dahlbeck, Douglas
 APPLICANT: Katagiri, Fumaki
 APPLICANT: Kunkel, Barbara N.

Qy	Neury March	3.4%;	Score 93.5;	DB 2;	Length 1461;
	Best Local Similarity	21.3%;	Pred. No. 4.4;		
	Matches	76;	Conservative	51;	Mismatches 116; Indels 111; Gaps 19;
Qy	105	TTEGTGLGAIIPRISSKYFN-----TTEDHKVEALRWASTHSPSWTLKICIDVO	157		
Db	531	TLDGT-----PRINIVFDFPPPLDVOYECEEVERY-----TEHGTP-----K	570		
Qy	158	PMQVYDLCOLLGKAEVQASFLKVMQDGRTPPSMRVLIDIVLEGSLSHIELONLT	217		
Db	571	PFKRDVA--FGEQSEDEF-----ENDLETDPPMOQLVSREVLLG-LKPCETIRQEV	623		
Qy	218	IDILVYDNHVAHARELK--GSFLRIYLSH-----TKLOSMNSENMQLSLEFHEHGTS	270		
Db	624	INELFYTERAHV-RLTKVLVDQVYQVRSREGILSPELRKIFSNEMLDQLHIGLH--E	679		
Qy	271	YGRGRVLPESNSDVQDLKKDLESANLT-----ANQHSVDYCQSEPPDSFPSSGSV	321		
Db	680	QMKAAVRKENET-SVIDDQIGED-----LTWFSQGEKILKHAATFCSSNP-----F	725		
Qy	322	SLYEVRCCOOLSATILTLTHOYLETPLCALKOKAPQOVRIRAKLSYKPRRLFQSVKHL	381		
Db	726	ALEMKRSQKQKDSRQTFVQDAESNPLCR-----RLQKLDIITQW-----	766		
Qy	382	CPKCHLLQAEVPHQGLDLDQDGAIRKTPDVKLQNTSLYDSKIMTTKNQGRKVAVH	437		

Query Match	3.3%	Score 91	DB 2	Length 425
Best Local Similarity	19.5%	Pred. No. 0.96		
Matches	92	Conservative	81	Mismatches 175; Indels 124; Gaps 22
Oy	23	WAVGVGVVFFKRR--YLSKGTD--YCSVVTIDQINVKITCL--LFSNSVWALPIIKYN	75	
Db	16	LSLSHGVDGSLDPDOLVNSKTSQSGFCNLLCGEIGIGKSTLMDLFTNFKFSDPTNHE	75	
Oy	76	GDIVFFHRLKIQIVYKKEGTOGIFSSGFASLTJFEGLTG-----ABIIPRTSKYFN	124	
Db	76	PGV-----RLKASYSLEQESNVR-----LKTITIDVTGFGDQIINKDSDYKPIVEYIAQFEA	127	
Oy	125	FTTEDHKWVEAL-----RWVASTH--MSPSWTLTKICDPQMPQVFDLTQLLGGKAEVDGA	177	
Db	128	YIQEELKIKRSLFNFNHDRIHACLYFIATGTGSLKSLDLVTYKKKLS-----	174	
Oy	178	SFLKVMGDTKTPPSWRVLIODLVLEGD---SHIHRQLNTLIDILVDNHVHARSUK	234	
Db	175	-----KVNIIPIIAKADTIIAKNELHAKRSKIMSELYSNG-----	208	

ATTORNEY/AGENT INFORMATION:
NAME: Henry, Janis C
REGISTRATION NUMBER: 34,347
REFERENCE/DOCKET NUMBER: 2631
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)470-4189
TELEFAX: (206)233-0644
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1569 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-458-791-2

Query Match 3.5% Score 94.5; DB 4; Length 1568;
Best Local Similarity 20.0%; Pred. No. 3.9;
Matches 99; Conservative 60; Mismatches 164; Indels 171; Gaps 26;

QY 91 KETGGIT-----SSGFSALTF-----EGTLCAPLIPRTSSKYENFTIEDHKMV 133
DB 208 KDTGRLATQELGRKLCBAGSLHFVDAFLMNGSIYFPY-----YNTS----- 255
QY 134 EALRWASTHMSPSWTLKLCDDVOPMOTFDLTCLLGAEDGASFLK-----VMD 185
DB 256 -----GAATGMPMARIAOSTEVLFOGASLDC---GHGHPDGRRLLSLSLVEALDYMA 307
QY 186 GT-----RTPPSMVRLIODLVLEGLSHIH-RLQNLITDILVYDNHV-----HV 229
DB 308 GVFSAAAGEGERSPTTALCLFR-----MSEIQARAKRVSMDFTAESHCKEGDOP 360
QY 230 ARSLKVSFLRIYSLHTRKLOSMNSENOTMLSLEFHLHGTSYGRGIRVLPESNEDVDOLK 289
DB 361 ERVOPIASSTLIHSDLSVGTVMNRTVFL-----GTGDG-----QLL 400
QY 290 KDESANLTANQHSADVICOSEPDSPSSGSVSLYEVEKCOLSATILTDHQLERTPL- 348
DB 401 KVIIGENLTSN-----C---PE-----VITEIK-----EETPVF 426
QY 349 -----CALIKOKAPOOYRIR-AKLRSYKP-RLLFOSVKLHCPKCHLLOEVPHEGD 396
DB 427 YKLVDPVKNIIYILTAGKEVRIRIVANCNKHKSCSECLTATDPHCGCHSLQCTFOGD 486
QY 397 -----LDIIFOGATTPPVKCLONTSLYDSKITTTNOKGRKVAHVFVANNGLP 446
DB 487 CVHSEMLENWLDI--SSGAKKCPKIQI-----IRSKKERTVTVMG-----SFS 529
QY 447 LSNCELLIEGTLSEICKLSNKFNSV-----IPVRSHEDELLELDLAPFLIOGTIHHYG 502
DB 530 RHSCGMK-NVDSRELCONKSOPIRRTCTCSIPTRATIKDVSVMVWFSGSNLSDRFN 588
QY 503 TGYCTP-----PICV 512
DB 589 FTNCSLKECPACV 602

RESULT 6
US-09-459-066-2
Sequence 2, Application US/09459066

GENERAL INFORMATION:
APPLICANT: Spriggs, Melanie
TITLE OF INVENTION: VIRAL ENCODED SEMAPHORIN PROTEIN
TITLE OF INVENTION: RECEPTOR DNA AND POLYPEPTIDES
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Janis C. Henry
STREET: 51 University St.
CITY: Seattle
STATE: WA
COUNTRY: US

ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS-DOS/Windows 95
SOFTWARE: Word for Windows 95, 7.0a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/459,066
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/958,598
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Henry, Janis C
REGISTRATION NUMBER: 34,347
REFERENCE/DOCKET NUMBER: 2631
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)470-4189
TELEFAX: (206)233-0644
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1568 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-459-066-2

Query Match 3.5% Score 94.5; DB 4; Length 1568;
Best Local Similarity 20.0%; Pred. No. 3.9;
Matches 99; Conservative 60; Mismatches 164; Indels 171; Gaps 26;

QY 91 KETGGIT-----SSGFSALTF-----EGTLCAPLIPRTSSKYENFTIEDHKMV 133
DB 208 KDTGRLATQELGRKLCBAGSLHFVDAFLMNGSIYFPY-----YNTS----- 255
QY 134 EALRWASTHMSPSWTLKLCDDVOPMOTFDLTCLLGAEDGASFLK-----VMD 185
DB 256 -----GAATGMPMARIAOSTEVLFOGASLDC---GHGHPDGRRLLSLSLVEALDYMA 307
QY 186 GT-----RTPPSMVRLIODLVLEGLSHIH-RLQNLITDILVYDNHV-----HV 229
DB 308 GVFSAAAGEGERSPTTALCLFR-----MSEIQARAKRVSMDFTAESHCKEGDOP 360
QY 230 ARSLKVSFLRIYSLHTRKLOSMNSENOTMLSLEFHLHGTSYGRGIRVLPESNEDVDOLK 289
DB 361 ERVOPIASSTLIHSDLSVGTVMNRTVFL-----GTGDG-----QLL 400
QY 290 KDESANLTANQHSADVICOSEPDSPSSGSVSLYEVEKCOLSATILTDHQLERTPL- 348
DB 401 KVIIGENLTSN-----C---PE-----VITEIK-----EETPVF 426
QY 349 -----CALIKOKAPOOYRIR-AKLRSYKP-RLLFOSVKLHCPKCHLLOEVPHEGD 396
DB 427 YKLVDPVKNIIYILTAGKEVRIRIVANCNKHKSCSECLTATDPHCGCHSLQCTFOGD 486
QY 397 -----LDIIFOGATTPPVKCLONTSLYDSKITTTNOKGRKVAHVFVANNGLP 446
DB 487 CVHSEMLENWLDI--SSGAKKCPKIQI-----IRSKKERTVTVMG-----SFS 529
QY 447 LSNCELLIEGTLSEICKLSNKFNSV-----IPVRSHEDELLELDLAPFLIOGTIHHYG 502
DB 530 RHSCGMK-NVDSRELCONKSOPIRRTCTCSIPTRATIKDVSVMVWFSGSNLSDRFN 588
QY 503 TGYCTP-----PICV 512
DB 589 FTNCSLKECPACV 602

RESULT 7
US-09-308-375-2
Sequence 2, Application US/09308375

```

Db 939 SNEYEKREIALKVTIHLNIGECNIGPOTG---EXCII-----EVNARLSRSA 990
Oy 62 FSGNTEALPIIYKNGDIVRFH---RLKIOVYKKEGOTITSSGFASLFEGLG--APIIP 116
Db 991 LASRATGPLAYISAKIALGDLISLKNSTTKTT-----ACEPESLDYITTKIP 1040
Oy 117 RTSSKYEFTT---EDHKWVEALRVMASTHMSPSWTLKLCVDQVPOGYDITCOLLGA 172
Db 1041 FMDLNKFEFASNTMNSKSVGEVMSIGRTFEESIQRCDIDNNYLGFSNTYC----- 1094
Oy 173 EVDGASFLKWDGTRTFPPSMRV--LIQDLVLEGDSLHHRLONLITDILYDNNHVA 230
Db 1095 -IDWDE--KRIIEELKNSPK-RIDAIHQAFHLMNPMOKIHELHIDWFL----- 1141
Oy 231 RSLKVGSLRIYSLHTRKLSMSENQMTLSLEFHLHGTSYGRGIRVLPESNSVDOLK 290
Db 1142 -----HKFYNIYNQNLKTKLLEQSLFNDLKYFKKHGFSKQIAHLYSFTSDNNNNN 1196
Oy 291 DLESANLTANOHSVDYICQSPDDSPSSGSVSLYEVEEQOLSATI--LIDHOYL----- 343
Db 1197 NISSCRVTEN---DVMKYREKLGLEP-----HIKVIDLSAEFPALTNYLYLYOQO 1245
Oy 344 ERTPLCALIKOKAPOOYIRAKLRBYKPRRLFQSVKLCPCCHLLQEVPHGEDLDIIFOD 403
Db 1246 EHDVLP LNKKR-----KICTLNNKRNANKKKVHV-KNHLVNEVVDKOTOLHKN 1295
Oy 404 GATKTPDKLQNTSLYDSKIWTKNQGRKVAHVFNKNGILPLSNECLLIEGGTISEI 463
Db 1296 NNNNN-----MNSGVNEKCKLNKESYGYNNSSNCINTNNINENNIC-----HDISINKN 1346
Oy 464 CK--LSNKFNSYIPVRSCHDELDLDSAPFLIOGTIIHYG 502
Db 1347 IKVTINNSNSTI-----SNNEVE-TNLNCVSEBAGSHHLYG 1382

```

RESULT 2

```

US-09-150-741-2
; Sequence 2, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
; APPLICANT: Stewart et al.
; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; Patent No. 6183996
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/150,741
; EARLIER FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PL6380
; EARLIER FILING DATE: 1992-12-16
; EARLIER APPLICATION NUMBER: AU93/00617
; EARLIER FILING DATE: 1993-12-02
; EARLIER APPLICATION NUMBER: 08/446,855
; EARLIER FILING DATE: 1995-07-06
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2391
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: protein
US-09-150-741-2

```

```

Query Match 3.98; Score 106; DB 4; Length 2351;
Best Local Similarity 20.98; Pred. No. 0.52;
Matches 109; Conservative 80; Mismatches 230; Indels 102; Gaps 25;

```

```

Oy 7 TINTIYPLNO--LKGGITVWYGV--VKFKPPYLSKGTIDYCSAVITVDQINXLI--CLL 51
Db 939 SNEYEKREIALKVTIHLNIGECNIGPOTG---EXCII-----EVNARLSRSA 990
Oy 62 FSGNTEALPIIYKNGDIVRFH---RLKIOVYKKEGOTITSSGFASLFEGLG--APIIP 116

```

```

Db 991 LASRATGPLAYISAKIALGDLISLKNSTTKTT-----ACEPESLDYITTKIP 1040
Oy 117 RTSSKYEFTT---EDHKWVEALRVMASTHMSPSWTLKLCVDQVPOGYDITCOLLGA 172
Db 1041 FMDLNKFEFASNTMNSKSVGEVMSIGRTFEESIQRCDIDNNYLGFSNTYC----- 1094
Oy 173 EVDGASFLKWDGTRTFPPSMRV--LIQDLVLEGDSLHHRLONLITDILYDNNHVA 230
Db 1095 -IDWDE--KRIIEELKNSPK-RIDAIHQAFHLMNPMOKIHELHIDWFL----- 1141
Oy 231 RSLKVGSLRIYSLHTRKLSMSENQMTLSLEFHLHGTSYGRGIRVLPESNSVDOLK 290
Db 1142 -----HKFYNIYNQNLKTKLLEQSLFNDLKYFKKHGFSKQIAHLYSFTSDNNNNN 1196
Oy 291 DLESANLTANOHSVDYICQSPDDSPSSGSVSLYEVEEQOLSATI--LIDHOYL----- 343
Db 1197 NISSCRVTEN---DVMKYREKLGLEP-----HIKVIDLSAEFPALTNYLYLYOQO 1245
Oy 344 ERTPLCALIKOKAPOOYIRAKLRBYKPRRLFQSVKLCPCCHLLQEVPHGEDLDIIFOD 403
Db 1246 EHDVLP LNKKR-----KICTLNNKRNANKKKVHV-KNHLVNEVVDKOTOLHKN 1295
Oy 404 GATKTPDKLQNTSLYDSKIWTKNQGRKVAHVFNKNGILPLSNECLLIEGGTISEI 463
Db 1296 NNNNN-----MNSGVNEKCKLNKESYGYNNSSNCINTNNINENNIC-----HDISINKN 1346
Oy 464 CK--LSNKFNSYIPVRSCHDELDLDSAPFLIOGTIIHYG 502
Db 1347 IKVTINNSNSTI-----SNNEVE-TNLNCVSEBAGSHHLYG 1382

```

RESULT 3

```

US-08-264-002-2
; Sequence 2, Application US/08264002
; Patent No. 5559019
; GENERAL INFORMATION:
; APPLICANT: GUI, JIAN-FANG
; APPLICANT: FU, XIANG-DONG
; TITLE OF INVENTION: NOVEL PROTEIN SERINE KINASE, SRPK1
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESS: SPENSLEY HORN JUBAS & LUBITZ
; STREET: 1880 Century Park East, Fifth Floor
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/264,002
; FILING DATE: 22-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: TUMARKIN PH.D., LISA A.
; REGISTRATION NUMBER: P-38,347
; REFERENCE/DOCKET NUMBER: PD3590
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/455-5100
; TELEFAX: 619/455-5110
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 655 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-264-002-2

```

ID NRPI_XENLA STANDARD: PRT: 928 AA.
AC P28824:
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 20-DEC-1992 (Rel. 24, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NEUROFILIN-1 PRECURSOR (A5 PROTEIN) (A5 ANTIGEN).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=91337458; PubMed=1908252;
RA Takagi S., Hirata T., Agata K., Mochii M., Eguchi G., Fujisawa H.,
RT "The A5 antigen, a candidate for the neuronal recognition molecule,
RT has homologies to complement components and coagulation factors.";
RL Neuron 7:295-307(1991).
CC -1- FUNCTION: RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE
CC CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF
CC CERTAIN NEURONAL CIRCUITS AND IN ORGANOREGULATION OF SEMAPHORINS (BY
CC SIMILARITY). IT MEDIATES THE CHEMOTACTIC ACTIVITY OF SEMAPHORINS (BY
CC SIMILARITY). PRESUMED TO BE INVOLVED IN THE NEURONAL RECOGNITION
CC BETWEEN THE OPTIC NERVE FIBERS AND THE VISUAL CENTERS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: RETINAL GANGLION CELLS AND VISUAL CENTER
CC NEURONS.
CC -1- SIMILARITY: BELONGS TO THE NEUROFILIN FAMILY.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D10467; BAA01260.1; -
DR InterPro: IPR000859; CUB.
DR InterPro: IPR000421; FA5B_C.
DR InterPro: IPR000998; MAM.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00754; F5_F8_Type_C; 2.
DR Pfam: PF00629; MAM; 1.
DR PRINTS: PR00020; MAMDOMAIN.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00231; FA5B; 2.
DR SMART: SM00137; MAM; 1.
DR PROSITE: PS00740; MAM_1; 1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01285; FA5B_C; 1; 2.
DR PROSITE: PS01285; FA5B_C; 2; 2.
DR PROSITE: PS00600; MAM_2; 1.
DR Transmembrane: Glycoprotein; Neurone; Signal; Repeat; Receptor;
RW Antigen.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 928 NEUROFILIN-1.
FT DOMAIN 22 860 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 861 883 POTENTIAL.
FT DOMAIN 884 928 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 27 141 CUB 1.
FT DOMAIN 147 265 CUB 2.
FT DOMAIN 275 424 F5/8 TYPE C 1.
FT DOMAIN 431 584 F5/8 TYPE C 2.
FT DOMAIN 646 812 MAM.
FT DISULFID 27 54 PROBABLE.
FT DISULFID 82 104 PROBABLE.
FT DISULFID 147 173 PROBABLE.
FT DISULFID 206 228 PROBABLE.

FT DISULFID 275 424 BY SIMILARITY.
FT DISULFID 431 584 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 150 150 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 523 523 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 844 844 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 928 AA: 103416 MW: 46032304.47890 CAC64;

Query Match 3.6%; Score 97.5; DB 1; Length 928;
Best Local Similarity 18.9%; Pred. No. 12;
Matches 91; Conservative 69; Mismatches 186; Indels 135; Gaps 22;

OY 29 VKFFKPPYLSKGDYCVATVYQTNVKKLCLFSGNVEALPIYK---NGDIVRRHAK 85
DB 355 VKSKYVDISSNGDWM---ITLKDKGNK---HLVFTGNTDATTDVYRPFKPVITRRVLR 407
OY 86 IOYKREKTCGTSSTSGFALTFE-----GTLGAPIIPTSSKYFNFT 126
DB 408 -----PVTWENGSLRFLYCGCKITDYPCSRMIGWVSGLISQITASSQVDVRRNV 458
OY 127 TEHKNVEALRWVA---STH-NSPSVTLKLCVOPMOTFDLTQOLGKAEDGASFL 181
DB 459 PELARLVTSRSGWALPSPSNTHPYTKEMQLDLAEKIVR---GYIIGGKHKEKVFEM 514
OY 182 KYMDGTRPPSPWVLI-----ODLVLEGLSHIRLQNTLIDLYVDNHYVARSLSKV 236
DB 515 KFKIGYSNKGTEPMIDSSKNKPKTEG-----NNYDPELRTFAHITT----- 560
OY 237 SFRLTY-----SLATKLOSMNSENOTMULF-----FLHGGTGYGRG 274
DB 561 GFIRIPEPASASGLALRLLELCEVEETPISPTPEVNGGDECEGLDLANCHSGTDEGFR 620
OY 275 IRV-LPESNSVDOLKKDLSEANLTANOSHVDVIOGSEPFDDSPSSGSVLYEVEPCQLS 333
DB 621 LTVGATGOSTETPTVEASPEEPDMT---HSDLDCKF---GWSQKRYKWMQNDISSDLK 673
OY 334 ATILT-----DHQYLERTPCALILKOKAPQOYRIAKLSYKPRFLFSV---KLH 381
DB 674 WAVLNSKITGPVQDH-----TGDGNIYSEADERHEGRH-----ARLMSPVYSSSRSAH 721
OY 382 CPRCHLQEVPHEGDDLIIFODGATKTPDYKLTQNTSLYDSKIWTNNOKG-----RRVAV 436
DB 722 CLTFWYHMDGSHVGTLSI-----KLKYEEDPFDQTLWTVSGNQGDKKREARYVL 771
OY 437 H 437
DB 772 H 772

Search completed: April 9, 2002, 17:11:19
Job time: 248 sec

CC CHANGE AND IS INVOLVED IN THE CHARGE-SELECTIVE ULTRAFILTRATION
 CC PROPERTIES. IT INTERACTS WITH OTHER BASEMENT MEMBRANE COMPONENTS
 CC SUCH AS LAMININ AND COLLAGEN TYPE IV AND SERVES AS AN ATTACHMENT
 CC SUBSTRATE FOR CELLS.
 CC -1 SUBUNIT: PURIFIED PERLECAN HAS A STRONG TENDENCY TO AGGREGATE IN
 CC DIMERS OR STELLATE STRUCTURES.
 CC -1 SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -1 TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES.
 CC -1 PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
 CC AND O-LINKED OLIGOSACCHARIDES.
 CC -1 SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
 CC -1 SIMILARITY: CONTAINS 10.5 LAMININ EGF-LIKE DOMAINS.
 CC -1 SIMILARITY: CONTAINS 22 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1 SIMILARITY: CONTAINS 3 LAMININ DOMAINS.
 CC -1 SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
 CC -1 SIMILARITY: CONTAINS 1 SEA DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M62515; CAA4373.1; -
 CC EMBL; M62529; AAA52700.1; -
 CC EMBL; M64283; AAA52699.1; -
 CC EMBL; S76436; AAB21121.2; -
 CC EMBL; L22078; -; NOT_ANNOTATED_CDS.
 CC HSSP; P00740; 11XA.
 CC MIM; 142461; -
 CC Interpro: IPR000561; EGF-like.
 CC Interpro: IPR000742; EGF_2.
 CC Interpro: IPR001438; EGF_11.
 CC Interpro: IPR003006; Ig_MHC.
 CC Interpro: IPR003598; Ig_C2.
 CC Interpro: IPR002172; LDL_recept-A.
 CC Interpro: IPR000034; Laminin_B.
 CC Interpro: IPR002049; Laminin_G.
 CC Interpro: IPR001791; Laminin_G.
 CC Interpro: IPR000082; SEA.
 CC Pfam: PF000008; EGF_4.
 CC Pfam: PF00047; Ig_22.
 CC Pfam: PF00052; Laminin_B_3.
 CC Pfam: PF00053; Laminin_EGF_8.
 CC Pfam: PF00054; Laminin_G_3.
 CC Pfam: PF00057; LDL_recept_a_4.
 CC PRINTS: PR00190; SEA; 1.
 CC PRINTS: PR0010; EGFLOOD.
 CC PRODOM: PD003031; Laminin_B_3.
 CC SMART: SM00180; EGF_Lam; 6.
 CC SMART: SM00001; EGF_Like; 8.
 CC SMART: SM00408; IGC2; 22.
 CC SMART: SM00281; Lamb; 3.
 CC SMART: SM00282; Lamb; 3.
 CC SMART: SM00192; LDLA; 4.
 CC SMART: SM00200; SEA; 1.
 CC PROSITE: PS00022; EGF_1; 9.
 CC PROSITE: PS01186; EGF_2; 5.
 CC PROSITE: PS01248; LAMININ_TYPE_EGF; 11.
 CC PROSITE: PS01209; LDLA_1; 4.
 CC PROSITE: PS50068; LDLA_2; 4.
 CC PROSITE: PS50024; SEA; 1.
 CC Signal: Basement membrane; Proteoglycan; Repeat; Glycoprotein;
 CC Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
 CC Extracellular matrix; EGF-like domain.
 CC SIGNAL 1 21 POTENTIAL.
 CC CHAIN 22 4393 BASEMENT MEMBRANE-SPECIFIC HEPARAN
 CC DOMAIN 80 194 SULFATE PROTEOGLYCAN CORE PROTEIN.
 CC DOMAIN 197 236 LDL-RECEPTOR CLASS A 1.

Query Match	Best Local Similarity	Score 102: DB 1: Length 4393:
FT DOMAIN	283	321
FT DOMAIN	323	361
FT DOMAIN	366	405
FT DOMAIN	405	506
FT DOMAIN	523	532
FT DOMAIN	533	732
FT DOMAIN	733	765
FT DOMAIN	766	815
FT DOMAIN	816	873
FT DOMAIN	881	925
FT DOMAIN	926	935
FT DOMAIN	936	1127
FT DOMAIN	1128	1160
FT DOMAIN	1161	1210
FT DOMAIN	1211	1267
FT DOMAIN	1277	1326
FT DOMAIN	1327	1336
FT DOMAIN	1337	1531
FT DOMAIN	1532	1564
FT DOMAIN	1565	1614
FT DOMAIN	1615	1672
FT DOMAIN	1679	1773
FT DOMAIN	1774	1867
FT DOMAIN	1868	1957
FT DOMAIN	1958	2053
FT DOMAIN	2054	2153
FT DOMAIN	2154	2246
FT DOMAIN	2247	2342
FT DOMAIN	2343	2438
FT DOMAIN	2439	2535
FT DOMAIN	2536	2631
FT DOMAIN	2632	2728
FT DOMAIN	2729	2828
FT DOMAIN	2829	2926
FT DOMAIN	2927	3023
FT DOMAIN	3024	3114
FT DOMAIN	3115	3213
FT DOMAIN	3214	3300
FT DOMAIN	3301	3401
FT DOMAIN	3402	3490
FT DOMAIN	3491	3576
FT DOMAIN	3577	3671
FT DOMAIN	3701	3847
FT DOMAIN	3846	3883
FT DOMAIN	3886	3924
FT DOMAIN	3966	4104
FT DOMAIN	4106	4143
FT DOMAIN	4145	4178
FT DOMAIN	4243	4391
FT SITE	65	67
FT SITE	71	73
FT SITE	76	78
FT SITE	4151	4153
FT SITE	4301	4303
FT DISULFID	199	212
FT DISULFID	206	225
FT DISULFID	219	234
FT DISULFID	285	297
FT DISULFID	292	310
FT DISULFID	304	319
FT DISULFID	325	337
FT DISULFID	332	350
FT DISULFID	344	359
FT DISULFID	368	381
FT DISULFID	375	394

Query Match 3.7%: Score 102: DB 1: Length 4393:
 Best Local Similarity 24.1%: Pred. No. 51:
 Matches 81: Conservative 38: Mismatches 137: Indels 80: Gaps 13:
 48 IIVDTNKLTLFLPSGNEALPIIYKNGDI-----VRFRLKI-----QVYKKE 92

DB 304 -REVAKV-TDDKSVK-----AALKODV-----SLSAVVLTEVKKHAGLPTHS 346
 OY 388 LQEPHEGDLDIIFODGATKTPDKLQNTSLYDKIWT-----TKNGGRKVA----- 435
 DB 347 LQDLHNDMDKEISSKOTFRTOYITREPDVKEWVSYDKRTKKPSHHGACAKGE 406
 OY 436 ----VHFVKNNGILPLSN---ECLLIEG-----TLSEICKUS 467
 DB 407 NIFQVQFLVKDASTOLNNNTYRVLLYTODGLGANFNVKAPDMLYKNNDARKKLEEVNELL 466
 OY 468 NKFN-----VIPVRS 479
 DB 467 TKFNSTYDAVVERRNG 482

RESULT: 5

GAT1_YEAST STANDARD: PRT: 510 AA.

AC P43574: 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE TRANSCRIPTIONAL REGULATORY PROTEIN GAT1.
 GN GAT1 OR YFL021W.
 OS Saccharomyces cerevisiae (baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CRP8;
 RX MEDLINE=96182087; PubMed=8622686;
 RA Coffman J.A., Rai R., Cunningham T., Svetlov V., Cooper T.G.;
 RT 'Gat1p, a GATA family protein whose production is sensitive to
 nitrogen catabolite repression, participates in transcriptional
 activation of nitrogen-catabolic genes in Saccharomyces cerevisiae.';
 RL Mol. Cell. Biol. 16:847-856(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RX MEDLINE=95400292; PubMed=7670463;
 RA Murakami Y., Naitou M., Hagiwara H., Shidata T., Ozawa M.,
 RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
 RA Yamazaki M., Tashiro H., Eki T.;
 RT 'Analysis of the nucleotide sequence of chromosome VI from
 Saccharomyces cerevisiae.';
 RL Nat. Genet. 10:261-268(1995).
 CC -1- FUNCTION: POSITIVE REGULATOR OF MULTIPLE NITROGEN CATABOLIC GENES.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: 027344; AAB03516.1; -
 DR EMBL: D50617; BAA09217.1; -
 DR HSSP: P17429; 5GAT.
 DR SGD: S0001873; GAT1.
 DR InterPro: IPR000679; ZNF_GATA.
 DR Pfam: PF00320; GATA.1.
 DR PRINTS: PR00619; GATAZNFINGER.
 DR SMART: SM00401; ZNF_GATA.1.
 DR PROSITE: PS00344; GATA_ZN_FINGER.1; 1.
 DR PROSITE: PS00114; GATA_ZN_FINGER.2; 1.
 KM Transcription regulation: DNA-binding; Zinc-finger; Nuclear protein.
 FT ZN_FING 310 334 GATA-TYPE.
 FT DOMAIN 151 158 POLY-ASP.
 SO SEQUENCE 510 AA: 56327 MW: 62080542695F35F CRC64;

Query Match 4.5%; Score 122; DB 1; Length 510;
 Best Local Similarity 17.9%; Pred. No. 0.065;
 Matches 80; Conservative 54; Mismatches 117; Indels 196; Gaps 15;

OY 209 HIRLQNLITIDLVYDNHV-HVARSKLVGSFLRIYSLHTKQSMNSGNQMLSEFHLHG 267
 DB 27 YIHCTRCYVNNITMTSTNRPNLDPLNLNK--EIMDLVSSAKKILPDSNRILNLSWRLHN 84
 OY 268 GTSYGRGIRVLPESNSDV----- 286
 DB 85 RTSFHRIRINIMQHSINDFSAPFASGVNAAGPCNNLDITDINOFPPLSDMLNNGSS 144
 OY 287 -----QLKDLSEANLTANQHS-----DVIGQSEPDSEF- 315
 DB 145 VEFNVFDDDDDDVDVETHSIYHSDLLNDMDSASQASHNAGCFPNFLDTSCSSSFDHFI 204
 OY 316 -----PSSGVSILVEVQCO---QLSATILIDHQYLEPTPLCA 350
 DB 205 FTNNLPFLNNNSINNHNSSHNNSPSIANNTNANTNTSASTNTNSPLLRNPS 264
 OY 351 ILKQAPQOYRIRAKLRYSYKPRPLFQSVYLHCPCHLQEPHEGDLDIIFODGAT---- 406
 DB 265 IYKPGSRNRSSVYKPKKPKALKIKSSSTV-----QSSATPPSN 301
 OY 407 --KTPDVRLQNTSLYDKIWTTRKQKRVAVHEVYKNGILPLNEC--LLIEGCT--L 460
 DB 302 TSSNPDIKSNCTTSTPLM-RKDPK-----LPLCNACGLFLKLHCVTRPL 347
 OY 461 S-----EICKLSKNFNSVIPVR 477
 DB 348 SLKTDIIRKRRSSRTKINNNTTPPPSSSLNPGACGKKKNTYASVASKRKSNIYAPLK 407
 OY 478 SCHEDELIDLSP---FLIGDTIHH 500
 DB 408 S--QDIPPKIASPSIPQYLRNSTRRHH 432

RESULT: 6

EX02_BP14 STANDARD: PRT: 560 AA.

AC P04522: 01-AUG-1987 (Rel. 05, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE EXONUCLEASE SUBUNIT 2 (EC 3.1.11.-) (PROTEIN GP46).
 GN Bacteriophage T4.
 OS Bacteriophage T4.
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae;
 OC T4-like phages.
 OX NCBI_TaxID=10665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85257446; PubMed=4018026;
 RA Gram H., Rueger W.;
 RT 'Genes 55, alpha gt, 47 and 46 of bacteriophage T4: the genomic
 RT organization as deduced by sequence analysis.';
 RL EMBO J. 4:257-264(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Kuter E., Aritsaka F., Kunisawa T., Tsugita A., Mosig G.,
 RA Mesyanzhinov V., Rueger W., Sliedman T., Thomas E.;
 RT 'Bacteriophage T4 genome analysis.';
 RL Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 478-560 FROM N.A.
 RA Hsu T., Karam J.;
 RL Submitted (APR-1987) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: EXONUCLEASE INVOLVED IN PHAGE DNA RECOMBINATION,
 CC REPLICATION, AND REPAIR.
 CC -1- SUBUNIT: CONSISTS OF TWO SUBUNITS, GP47 AND GP46.
 CC -1- SIMILARITY: STRONG TO T5 PROTEIN D13 AND TO YEAST RAD52.

Db 69 VNVVFSQNFEDLPIIQRWDIVRHARLQHYNDAKOLNVNRYSSMCLFIGNDKEAP 128
 Oy 114 IIPRT-----SKRYF-----NETTEDI--KNVEALRVASTHMSPTL--LKLC 154
 Db 129 LERKVENEDOTNNYFTPTPNFSGKSTGTQGHCKILKDKKSKDYFSNDVVEQYKKA 188
 Oy 155 DVOPMOT---PULTCOLLGAKEVDGASFLKKWDGT-----RTPPSRFVLIODL 201
 Db 189 DIETAMKNTKIDFLAKKVEISDNDYTNVSLNDSTGTWIGHLKRFKPPH----- 240
 Oy 202 VLEGDSLHRIHLONLITDILVYDNHVARSLKVSFLRLYSLHTKLOSNSEN 255
 Db 241 LVAGDVLRITSVAKEDNSLIFSSSHNLK-----FFSSSIHKKLKSSSD 288

RESULT 2
 TEBA_EUPCR STANDARD: PRT: 460 AA.
 AC 006183;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DE TELOMERE-BINDING PROTEIN HOMOLOG.
 OS Euplotides crassus.
 OC Eukaryota: Alveolata; Ciliophora; hypotrichs; Euplotidae; Euplotidae;
 OC Euplotidae.
 OX NCBI_TaxID=5936;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-93126105; PubMed-1480483;
 RA Wang M., Skopp R., Scofield M., Price C.;
 RT Euplotides crassus has genes encoding telomere-binding proteins and
 RT telomere-binding protein homologs.*;
 RL Nucleic Acids Res. 20:6621-6629(1992).
 CC -1- FUNCTION: MAY BIND TELOMERIC TAGC SEQUENCES.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -1- SIMILARITY: BELONGS TO THE TELOMERE-BINDING PROTEIN ALPHA SUBUNIT
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: M96819; AAA29128.1; -
 DR PIR: S35525;
 DR InterPro: IPR003415; Telo_bind_alpha.
 DR Pfam: PF02307; Telo_bind_alpha; 1.
 DR KJ: DNA-binding; Nuclear protein; Telomere; Multigene family.
 SQ SEQUENCE 460 AA: 53360 MW: EDIC141385A0B5FE CRC64;

Query Match 5.7%; Score 156.5; DB 1; Length 460;
 Best Local Similarity 21.4%; Pred. No. 0.00012;
 Matches 116; Conservative 74; Mismatches 196; Indels 155; Gaps 24;

Oy 9 YITPLNQL--KGGIVNVGVKFRPPYLSKGTDCSVVITVDOT--NFK-----LTC 59
 Db 19 YETIEIGSIEEENASINFAVVIDACFPYKVDKRYWCLKVIDITHNKEGDNDFAIV 78
 Oy 60 LIFSAGTEALPIIYKNDIYFRHLKIOYKKETOGT-----SSGASLIFETGLCAP 114
 Db 79 ALQSRKEDLPIDQCDIIRHRAEYN--YKDDHYFKLNKSYSSSMALFSADEVAPEV 137
 Oy 115 IP-----RTSSKYFETEDHKVNEALVMASTHMSPTLTKLCDVQPMQYFD 163
 Db 138 IKDGDDEFTYRSYVSGKYNFOTQDOKLKNTRAMKNSYFAKN-----DV----- 183
 Oy 164 LTCOLLRAEYDASFLKKWDGTRTPPSRVRVLIODLVLEGDSLHRIHLONLITDILVY 223

Db 184 -----IDEMTTPLSQARQEBGDENVGKVTQIYHRDYTSDLRK 224
 Oy 224 DN-----HVVHVARSLKVGSLRLY--SLHTKLOSMSSENQTMLSLFFHLHGGTSGRGIRV 277
 Db 225 DISKATMFLTVSR-----KFRPLIEGVIIKISVINSD-----TERRCLEL 268
 Oy 278 LPESN---SDVDOLKKDLESANLTANOHD-----VICQSPDDSPSSGSVSLYEYER 328
 Db 269 APMNSMTFVPSRLAKSDQISLSPDKVDKELIKVLTLP----- 311
 Oy 329 COOLSATILTDHOYLERTPICAILKOKAPO--YRIRAKLRSPKRRLLFOSYKLCPCKH 386
 Db 312 --VLATTTGGDSELPLTLESEFEDYVDKDAVFARFSLIKITDRVEDYVEETPR-- 367
 Oy 387 LLOEVPHEGDDLIIFODGATKT--PDVKLO-----NTSLYDS--KIWT--TKNQGRKYAV 436
 Db 368 -----GAPRSKPYKVOFLKIDPSTALNDMLYKITYLSHDLKKEFF- 409
 Oy 437 HEVKNNGILPLSNECLLIEGTLSEICLSKMFNSVIPRSGHDELIDLSAPFLIOG 496
 Db 410 -----PGVDPSSAQ-----TPSGHSLT-RKYASTLMKRVNHDVLEKGAFFIRD 455
 Oy 497 T 497
 Db 456 T 456

RESULT 3
 TEBA_OXYNO STANDARD: PRT: 495 AA.
 AC P29549;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE TELOMERE-BINDING PROTEIN ALPHA SUBUNIT (TELOMERE-BINDING PROTEIN 56
 DE KDA SUBUNIT) (TEBA ALPHA).
 GN MAC-56A AND MAC-56K AND MAC-56S.
 OS Oxytricha nova.
 OC Eukaryota: Alveolata; Ciliophora; hypotrichs; Stichotrichida;
 OC Oxytrichidae; Oxytricha.
 OX NCBI_TaxID=5945;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE-92035001; PubMed-1840510;
 RA Gray J.T., Celander D.W., Price C.M., Cech T.R.;
 RT "Cloning and expression of genes for the Oxytricha telomere-binding
 RT protein: specific subunit interactions in the telomeric complex.*";
 RL Cell 67:807-814(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92249771; PubMed-1577273;
 RA Mitcham J.L., Lynn A.J., Prescott D.M.;
 RT "Analysis of a scrambled gene: the gene encoding
 RT alpha-telomere-binding protein in Oxytricha nova.*";
 RL Genes Dev. 6:788-800(1992).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
 RX MEDLINE-99091054; PubMed-9875850.
 RA Horvath M.P., Schweikert V.L., Bevilacqua J.M., Ruggles J.A.,
 RA Schultz S.C.;
 RT "Crystal structure of the Oxytricha nova telomere end binding protein
 RT complexed with single strand DNA.*";
 RL Cell 95:963-974(1998).
 CC -1- FUNCTION: MAY FUNCTION AS PROTECTIVE CAPPING OF THE SINGLE-
 CC STRANDED TELOMERIC OVERHANG. MAY ALSO PARTICIPATE IN TELOMERE-
 CC LENGTH REGULATION DURING DNA REPLICATION. BIND SPECIFICALLY TO THE
 CC 14G4-CONTAINING EXTENSION ON THE 3' STRAND AND PROTECT THIS REGION
 CC OF THE TELOMERE FROM NUCLEASE DIGESTION AND CHEMICAL MODIFICATION.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- MISCELLANEOUS: THE SEQUENCE OF THE A (OR ALANINE) VERSION IS
 CC SHOWN. THE S (OR SERINE) VERSION DIFFERS IN ONLY TWO POSITIONS,
 CC AND THE K VERSION IN ONLY ONE. THERE MAY BE OTHER VERSIONS IN

OY 164 -LTCOL-IGKAEVDGASFLKVMGDTGTRTPPSRWLVLO----DLVLEGD----LSHRL 213
DB 228 KFSCHIRVGNPKTKLRSTTVYKFAKEIP-----VIVENNSTVYVERPTCLKLVNFKR 282
OY 214 ONLT--ID-ILVYDNH--VHVARSLAVG--SFLRYSLSHTKLOS---MNSENOIMLSLEF 263
DB 283 ANITWTFIDCSFLHDEKEGIYITNEERKKGKDFELKSVLIRVSHNKPASDNLTIMCM-- 340
OY 264 HLHGSTYSGRGIRVLPESNDVD-OLKKOLESAN--LIANOHSVDYICQSPDPS-----F 315
DB 341 ----ALSPYGNKVMNISSEKITFLGSEISSTDPUSVTE-STLDTQSPASSVSPARY 395
OY 316 PSSGSVSLYEV 326
DB 396 PATSVTLVDV 406

RESULT 12
T00342
hypothetical protein KIA0580 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00342
R:Magase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
DNA Res. 5, 31-39, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complete
A:Reference number: 214086; MUID:98290545
A:Accession: T00342
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1044 <NAC>
A:Cross-References: EMBL:AB011152; NID:93043683; PIDN:BAA25506.1; PID:93043684
A:Experimental source: brain; clone HJ0601
C:Genetics:
A:Note: KIA0580

Query Match 4.0%; Score 110.5; DB 2; Length 1044;
Best Local Similarity 20.2%; Pred. No. 3.4;
Matches 102; Conservative 68; Mismatches 171; Indels 143; Gaps 26;

OY 49 IYDQTVNKLTLCL--FSGNYEALPIIYKNGDIVRFRHLKIOYKKETOCITSSGASLT 105
DB 255 LLEETN-KKMCVLEGGFLSTYENDKSTTPNGTININEVICLAHKEDFYLTGPIF--I 310
OY 106 FEGTLCAPLIIPRTSYFNTTEDHKVLEALRYMA--STHMSPTLKLCDVQPMQYF 162
DB 311 FE-----IYLPSEVFLFGAETS--QOQRKMTETIAKHFPPLA-ENLTAD----Y 355
OY 163 DLTCOLGKAEVDGASFLKVM--DGTPTPFSRWLVLODLVLEGDLSHRLONLTID 219
DB 356 DLTCOLGKAEVDGASFLKVM--DGTPTPFSRWLVLODLVLEGDLSHRLONLTID 219
OY 220 ILVYDNH--VHVARSLAVG--SFLRYSLSHTKLOS---MNSENOIMLSLEF 263
DB 410 TMQONEKEKLDVLLVENGKRLTYIHG-HTKLDTVMTALFEKAAGTGNALODQOLSKNDV 468
OY 258 MSLFELHNGTSGRGIRVLPESNSD-----VDLKKDLESANLIANOH--SDVTC- 307
DB 469 PIVNSCIAFVTOYGLGCKITYKNGDPLHISELSFKKADARSFKLRACKHOLEDTAV 528
OY 308 ----GSEPDPSFGSGSVSLY-----EVERCOOLSATILTDHOYLETPICAIL 352
DB 529 LKSFSLDIDALITK--ELYPYISALDITODDERAKIKKYG-A-FIRSLPGVNRATIAAI 584
OY 353 KQAKAQQOYRIARAKLSYKPRRLFOYVKLHCKHQLQEVNHEG--DLDIDFQGGATKTPD 410
DB 585 E-----ELRYO-----KC-----SEINMNAHNAHALV----- 608
OY 411 VKLQNTSLYDSKIMITKNQGRKVAVHFVKNNGILPLSNCECLLILLEGTLSEICKLSNKF 470
DB 609 -----SSCLPOTKGTGOTSEEVNVIEDLI 630

OY 471 NSVIPVRSGHED-LELJLISAPFL 493
DB 631 NNTVEIFEVKEDQVKOMDIENSFTI 654

RESULT 13
T21403
hypothetical protein F26D2.10 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C:Accession: T21403
R:McMurtry, A.
submitted to the EMBL Data Library, November 1996
A:Reference number: 219418
A:Accession: T21403
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1217 <MIL>
A:Cross-References: EMBL:281513; PIDN:CAB04180.1; GSPDB:GN00023; CESP:F26D2.10
A:Experimental source: clone F26D2
C:Genetics:
A:Gene: CESP:F26D2.10
A:Map position: 5
A:Intons: 29/1; 568/2; 608/3; 746/1; 871/3; 910/1; 985/1; 1018/3; 1124/1; 1170/3
C:Superfamily: Caenorhabditis elegans hypothetical protein F36D3.5

Query Match 4.0%; Score 108; DB 2; Length 1217;
Best Local Similarity 20.6%; Pred. No. 6.7; Mismatches 166; Indels 172; Gaps 30;

OY 45 SVYTIYDQTVNKLTLCL--FSGNYEALPIIYKNGDIVRFRHLKIOYKK 91
DB 483 AYSVTLDOVNANSISIIYKRLIRRYKGIYVNSYFELSTIKIARIAS----ELQVQK 538
OY 92 ETQGITSSGASLTGEGTLCAPLIIPRTS-----SKYFNTTEDHKVLEALRYMASTH 143
DB 539 TAENITVS-----DGNLTGDLISMLSVETSKCPSOYKTPMPTLESTVSKLK--SLR 588
OY 144 MSPSKTLKLCDVQPMQYFDLTGOLGKAEVDGASFLKVMGDT-----RTP 190
DB 589 MFPDGETISGM--LEID--LLGKYOSNLTLVEKEKSTSHVNSTNPPVLLKTP 640
OY 151 FFSRWLVLODL--VLEGDLSH--HR-----LQNLITD--ILVYDNH----- 228
DB 641 ---QALINIGTSTVLE-DLENVRQHQFLYSIRNFTADYKVIYGENLLEIMIDPEK 695
OY 229 -----VARSIAVGS--FLRYSLSHTKQSNMSENOTLSLEFL--HGCT- 269
DB 696 LRELEBADQNGFASGLRNGSLTEMARVPEDEAKITGIGGDKSMANVSINLLQASGTD 755
OY 270 --SYGRGIRVLPESNDVDOLKKDLESANLT-----ANOHSVDI--QSEPDPS 314
DB 756 QLAASVFKIRIELANDPFAKHEASLSKASAVTYSLSLAYDEIFGNSKSKILTEKIDNS 815
OY 315 FP-----SSGSV-----SLY-----EVERCOOLSATIL 337
DB 816 FPMIYIILLICIGISILLIGFCVAVGYFTNGRAKYTNLLYYFGKPLDERRRRYS--F 873
OY 338 TDHOYLETPICA--ILKQAKAQQOYRIARAKLSYKPRRLFOYVKLHCKHCHLOE-- 390
DB 874 LQNOGDKNTLLDVREIKNTINWKAVALGSIYITAYK--FGTALMAATKSALPEIYKI 930
OY 391 -VPHSGDLDIIFQDGATKTPDVQLQNTSLYDSKIMITKNQGRKVAVHFVKNNG 443
DB 931 LIEHGADLTLL--NTLNKTPPEQMIPLDSPDPSDKITLKKIKRLK--YKNKANG 979

RESULT 14
T18410
carbamoyl-phosphate synthase (glutamine-hydrolyzing) (EC 6.3.5.5) II - malaria parasi
C:Species: Plasmodium falciparum

Db 345 ELKQTSVNEITAHVINEPASIKLITITLHAPLQNLKPKRHRLRVQVVDMPKSLQF 404
 Oy 378 VKLHCPKCH-----LLOEVPHEGDLIIF--QDQATKTPDVKLOMTSLD 420
 Db 405 AVLSQPPSSVYVMFALLVROVSNV-TLPVILFFSDAELINSSKTIQPCNLAD 455

RESULT 2

S35524
 telomere-binding protein, 51k - Euplotes crassus

C:Species: Euplotes crassus
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 07-Dec-1999
 C:Accession: S35524
 R:Wang, W.; Skopp, R.; Scofield, M.; Price, C.
 Nucleic Acids Res. 20, 6621-6629, 1992
 A:Title: Euplotes crassus has genes encoding telomere-binding proteins and telomere-binding proteins
 A:Reference number: S35524; MUID:93126105
 A:Accession: S35524
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-446 <MAN>
 A:Cross-references: EMBL:M96818; NID:g159023; PID:g159024
 C:Genetics:
 A:Genetic code: SGC9
 A:Introns: 9/3
 C:Keywords: DNA binding; nucleus

Query Match 6.0% Score 164; DB 2; Length 446;
 Best Local Similarity 23.8%; Pred. No. 7.2e-05;

Matches 70; Conservative 55; Mismatches 107; Indels 62; Gaps 13;

Oy 8 NYITPLNQLK--GGITVNVYGVKFFKPPYLSKGTID-YCSVVTIVDQTNV-----K 56
 Db 11 HGYQSDLSIKKEGEEDQYHEGVVIDASEFY--KGEKRYVTCTKVADPSSVAKGKRLNT 68
 Oy 57 LTCLFSGNEALPIIKKNDIVAFHRLKIOVKKETGQITSSGFAS--LTFEGTLGAP 113
 Db 69 VNVVFFSONEDLPILIRVGDIVVHRAQLQHTNDAKQLNVNMYRSSKCLFIGNDKREAP 128
 Oy 114 IIPRT-----SSKYF-----NFTEDH--KRVKALRVMASTMSPSMTL--LRLC 154
 Db 129 LEPEVNEEDGTNNFYSTPYNFSKGSFTQEGHETKILKDKMSKDYFSNNDDVEGYKKA 188
 Oy 155 DVQPMQY----FDLTQQLKAEVDGASFLKLVMDGT-----RTPPSMVLVLDL 201
 Db 189 DIETAMKNTKIDELAKVTEISDNDQYITNTVSLNDSTGQTWTGHLFRKRPVH----- 240
 Oy 202 VLBDLSHIRLQMLTIDILVYDNHVVARSLSKVSFLRIYSLHTKQSMNSN 255
 Db 241 LVKGDVLRKSVSAKEDNSLIFSSHNLK-----FFSSSIHKKLKSSISSD 288

RESULT 3

S35525
 telomere-binding protein - Euplotes crassus

C:Species: Euplotes crassus
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 07-Dec-1999
 C:Accession: S35525
 R:Wang, W.; Skopp, R.; Scofield, M.; Price, C.
 Nucleic Acids Res. 20, 6621-6629, 1992
 A:Title: Euplotes crassus has genes encoding telomere-binding proteins and telomere-binding proteins
 A:Reference number: S35524; MUID:93126105
 A:Accession: S35525
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-460 <MAN>
 A:Cross-references: EMBL:M96819; NID:g159025; PID:g159026
 C:Genetics:
 A:Genetic code: SGC9
 A:Introns: 18/3; 377/3
 C:Keywords: DNA binding; nucleus

Query Match 5.7% Score 156.5; DB 2; Length 460;
 Best Local Similarity 21.4%; Pred. No. 0.00029;
 Matches 116; Conservative 74; Mismatches 196; Indels 155; Gaps 24;

Oy 9 YIYPLNQL--KGGITVNVYGVKFFKPPYLSKGTIDYCSVVTIVDQTNV-----LTC 59
 Db 19 YETIEIGSIEENENASINFVAVYIDACFPYKVDKRYKMYLKVITDTHNVKEDDNFALV 78
 Oy 60 LTFSGNEALPIIKKNDIVAFHRLKIOVKKETGQIT-----SSGFASLTFEGTLGAP 114
 Db 79 ALDSRKFEDLPILIRVGDIVVHRAQLQHTNDAKQLNVNMYRSSKCLFIGNDKREAP 137
 Oy 115 IP-----RTSSKYFNEFTEDHKNVKEALRVMASTMSPSMTLKLCDVQPMQYED 163
 Db 138 IMKEGDOFTFRSVAYSGSKOYVFDQDQKLKLNRAHMKSYFAKN-----DV----- 183
 Oy 164 LTCQLKAEVDGASFLKLVMDGTTRPPSPMRYLIDLVLEGLSHIRLQNLTDILVY 223
 Db 184 -----IIDENYTPLSQARQEEGDFNVGVKVTQIVHRDYTTSOLRVK 224
 Oy 224 DN-----HVVARSLSKVSFLRIY-SLHTKLSQSMNSQNTMLSLFHLHGTSYGRGIV 277
 Db 225 DTSKATVFLIVSR--KPRLEGVYIKIRSVNIDSE-----TERECLEL 268
 Oy 278 LPESN---SDVDQLKDLASMLTANQHSN-----VICSEPDSPSSGSVSLYEVR 328
 Db 269 APHSNIMTFVFPFSLKSLDSQISLSPDKVDKELIKKVLITP----- 311
 Oy 329 COQLSATILIDHOYLEKTPICAILKOKAPOO--YRIFAKRSKPPRLQSVKLBKPKCH 386
 Db 312 --VLAITTFDYSLEPLTEISELFEVDYTDKDAVFAFSLTKITTPRVEVEYEEYPR-- 367
 Oy 387 LLOEVPHEGDLIIIFQDQATKT--PDVRLQ-----NTSLDS--KIMT--TKNKGKRVAV 436
 Db 368 -----GARRSKPVYKVQFLINDPSALNDNLYKLYLSHGLGKEFF-- 409
 Oy 437 HFVKNQCIPLSNECLLIGGLISEICKLSKNFNSYIPRSHEDLELDLSAPPLIG 496
 Db 410 -----PCVDPSAQQ-----TPSGHSL--RRYASTLKNFNHDAVLEKGAFFIRD 455
 Oy 497 T 497
 Db 456 T 456

RESULT 4

A41221
 macronuclear alpha protein (alanine version) - Oxytricha nova

C:Species: Oxytricha nova
 C:Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 02-Mar-2001
 C:Accession: A41221; B41221
 R:Gray, J.T.; Celander, D.W.; Price, C.M.; Cech, T.R.
 Cell 67, 807-814, 1991
 A:Title: Cloning and expression of genes for the Oxytricha telomere-binding protein:
 A:Reference number: A41221; MUID:92035001
 A:Accession: A41221
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-495 <GRA>
 A:Cross-references: GB:M68931; NID:g159811; PID:g159812
 A:Accession: B41221
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-310,'S',312-455,'E',457-495 <GR2>
 A:Cross-references: GB:M68930; NID:g159813; PID:g159814
 C:Genetics:
 A:Genetic code: SGC5
 C:Keywords: DNA binding; nucleus

Query Match 5.4% Score 148.5; DB 2; Length 495;
 Best Local Similarity 19.6%; Pred. No. 0.0013;

RESULT	11
PFMAL13P2_1	
WPCOMMENT	

gene /dev_stage="adult"
/tissue_type="uterus"
/map="530.1 CR from top of Chr7 linkage group"
1. 411
/gene="DKFZ5860211"

CDS
/gene="DKFZ5860211"
/gene="DKFZ5860211"
/note="WUGSC:H.DJ0907C10.1, partly"
/codon_start=1
/product="hypothetical protein"
/protein_id="CA043281.1"
/db_xref="GI:4884146"
/translation="DHGKCKSSLSRLSIONLSYDKTSMISSVAELGIVLPLOYV
VATFLDDGIVLEAVLMDSDKFFQIPASEVLMDDDLQKSDVMIMDMCPSPKIDAY
PWLECFIKSYNTNGTDNQCIVQIDTIVAEVY"
polyA_signal
polyA_site 1983
BASE COUNT 631 a 343 c 357 g 681 t
ORIGIN

Query Match 17.1%; Score 310.4; DB 9; Length 2012;
Best Local Similarity 99.7%; Pred. No. 3.9e-64;
Matches 311; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1505 gcactggatgttaccctccaatagtgtgtgtatagacccttaccattgatatga 1564
DB 100 GCACCTGGTATTGTACCCCTCCCAATAGTGTGTATGACCTTACACTTGATGATGA 159
OY 1565 acaggaagtactagaagcctatcctcagatgattcgacaattcttcagattccagatca 1624
DB 160 ACAGGAGTACTACAGACCTATCTCATGATTCGACAAATCTTCCAGATTCAGCATCG 219
OY 1625 gaagttctgtatgatatgaccttcagaaaagtgtgatalgatalgatalgttctg 1684
DB 220 GAAGTCTGTGATGATGATGACCTTCAGAAAGTGTGATGATGATGATGATGATGAT 279
OY 1685 cctcccggaataaattgatatcctggtgtgtgaaatgcttcacatcaagtatacaat 1744
DB 280 CCTCCAGGATTAATAATGATGATATCCCTGTTGAAAGTTCATCAAGTATCAAT 339
OY 1745 gtccacaatggaacataatcaattgctatcagattttaccacacagttcgagaa 1804
DB 340 GTCACAATAGCAACATATCAAAATTGCTATCAGATTTTTCACACACAGTTGCAGAA 399
OY 1805 gatgtaactaa 1816
DB 400 GATGTAATCTAA 411

RESULT 8
AC004925 154959 bp DNA PRI 21-DEC-1999
LOCUS Homo sapiens PAC clone RP5-907C10 from 7q31-3q32, complete
DEFINITION
AC004925
AC004925
AC004925.1 GI:4156174
KEYWORDS
HITG.
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
1 (bases 1 to 154959)
Jones, R., Tin-Mollam, A. and Keppeler, D.
TITLE
The sequence of Homo sapiens PAC clone RP5-907C10
JOURNAL
Unpublished
2 (bases 1 to 154959)
REFERENCE
AUTHORS
Waterston, R.H.
TITLE
Direct Submission
JOURNAL
Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
3 (bases 1 to 154959)

AUTHORS
Waterston, R.H.
TITLE
Direct Submission
JOURNAL
Submitted (14-JAN-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
4 (bases 1 to 154959)
Waterston, R.
TITLE
Direct Submission
JOURNAL
Submitted (22-JAN-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
5 (bases 1 to 154959)
Waterston, R.
TITLE
Direct Submission
JOURNAL
Submitted (21-DEC-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jan 14, 1999 this sequence version replaced gi:3213072.
COMMENT
Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: saplense@wustl.wustl.edu
Summary Statistics
Center project name: H.DJ0907C10

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone; and the assembly was
confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
<http://www.nih.gov/DIR/STB/CHR/>, send
<mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
This clone was derived from human PAC library RPI-5, prepared by
Pieter de Jong and coworkers at the Roswell Park Cancer Institute
(<http://bacpac.med.buffalo.edu>) using the method described by
Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from
one male donor.
The clone may be obtained either from Genome Systems, Inc.
(<http://www.genomesystems.com>) or Research Genetics, Inc.
(<http://www.resgen.com>); or from Pieter de Jong.
VECTOR: pCYPAC2
NEIGHBORING SEQUENCE INFORMATION:
Actual start of this clone is at base position 1 of RP5-907C10;
actual end is at 154959 of RP5-907C10.

The sequence from 4689 to 5920 is from PCR product from PAC
RP5-907C10 DNA.

FEATURES
source
Location/Qualifiers
1..154959
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q31-3q32"
/clone="RP5-907C10"
/clone_lib="RPI-5"
66..259
/rpt_family="L1"
repeat_region

Db 869 TTTCGATCTTTGACGTTTGAGGGAAGCTTTGGAGCCCTTATCATACCTCGCACTTCAAGC 928

Qy 361 aagatatttaactcaactcaactgagagccacaanaatgtagaagccttaactggtttggca 420

Db 929 AAGTATTTTAAGTCACTACTGAGGACCAACAAATGTAAGAACTTACTGTTTGGCA 988

Qy 421 tctaactcaatgtaacccgctctggacataactaanaatgtgtgagttcagccaatgca 480

Db 989 TCTACCTCATATGTCACCGCTTGAGCATTTACTAAATGTGTGATGTGACCAATGACAG 1048

Qy 481 taattgacctactgtaactgctctggccaagaagtagaggaacatcatcttct 540

Db 1049 TATTTGAGCTGACTGCTGTACCTCTTGCGCAAGACAAAGATGATGGACATCATATCTT 1108

Qy 541 ctaaaagtagagtaggagccagagacacatttccaactctggagagcttaataacaagac 600

Db 1109 CTAAAGGTATGGATGGACAGACAGACACCATTTCCATCTTGAGAGCTTTAATACAGAC 1168

Qy 601 ctgtgtcttgaaggtgatttaagtagcaatccatccgctcacaanaatctgacaatagac 660

Db 1169 CTGTGCTTGAAGGTATTTAAGTCATCTCATTCGCTACAAATCTGACAAATAGACAT 1228

Qy 661 ttagtcaagataacacatgltcaactgtagcaagaactctgaaggttggaagcttctaga 720

Db 1229 TTAGCTACGATTAACCATGTTCTATGTGGCAAGATCTCTGAGGTTGGAACTTTCTTACA 1288

Qy 721 atctatagccttcaacaaactccaactcaatgaatcagaagaatcagaacatgtaagt 780

Db 1289 ATCTATAGCCTTCATACCAAACTTCAATCATGATGATGAGCAATGAGCAAAATGTTAAGT 1348

Qy 781 ttagagttcatcttcaactgaggttagcaagtagttagcaggttagggtctgtccaga 840

Db 1349 TTAGATTTTCACTCTTATGAGAGGTACCAAGTTATGTCGGGAATTCAGGCTTTGGCAGAA 1408

Qy 841 agtaactctgtagtggatcaactgtaaaagatttagaactctgcaacttgacaagccaat 900

Db 1409 ACTAACTTGATGTGATGATCACTGAAAGAAAGATTAGAAATCTGCCAATTTGACAGCTAT 1468

Qy 901 cagcatcagatggttactctgtcaactagaacctgacagagcttccaagctctgtagta 960

Db 1469 CACCATTCAGATGTTATCTGTCAATGAGAACCTGATGACAGCTTTCCAAAGCTCTGATGA 1528

Qy 961 gtaacttaacagagtagaagaatgltcaagagctatctgtctacataacttaacatcat 1020

Db 1529 GATATCTTGTATGAGGTAGAAAGATGTCACACCTATCTGTCTACATTTACAGATCAT 1588

Qy 1021 cagtaacttgagagagacacacactatgtgcatcttgaacaaagaagctcctcaacatac 1080

Db 1589 CATATTTTGGAGAGGACACACTGTGTGCATTTTGAAACAAAGCTCTCTCAACATAT 1648

Qy 1081 cgcattccgagcaaaatgtaggtcatactaaagccagaagaactatctagtcgttaact 1140

Db 1649 CGCATCCGAGCAAAATTTAGGTCATTAAGCCCAAGACATTTTCAGTCTGTTAACTT 1708

Qy 1141 catlgtcccaaatgtcatctgtctgaagaagttccacatgaaggaggttggatataat 1200

Db 1709 CATTGCCCTAAATGTCATTTACTGCAAGAAGTTCCAGCTGAGGGTATTTGGATATATTT 1768

Qy 1201 tctcagagtagtgcacaataaaccagatgltcaagctacaataacatcatatagat 1260

Db 1769 TTGCAGAGTAGTGTCAAACTCCAGATGTCAACTTCAAATATACGACATTTATGCAAT 1828

Qy 1261 tcaaaaactgagcacacttaaaaactcaaaaagagcgaagaagtagcggttcatcttgtgaa 1320

Db 1829 TCAAAAATTTGAGCACTTAAATATCGAAGAGCGAAAGTAGCACTTATTTTGGAAA 1888

Qy 1321 aataatgtagtctccgcttccaatgtagtctactcttgaagaagtagtaaac 1380

Db 1889 AATTAATGATATCTCCCACTTTCAATATGATGCTACTTTTATAGAACAGAGACTACACTC 1948

Qy 1381 agtgaataattgcaaacctctgacaagaagtttaagtagtaactctgtgtgagcttgccac 1440

Db 1949 AGTAATTTGCAAACTCTGAAACAGTTTAATAGTAAATCTGTGAGATCTGGCCAC 2008

Qy 1441 gaagaccgtgaaacttltgaccttcaagcaccattcttatacaagaagaatacatcac 1500

Db 2009 GAAGACCTGGAACCTTTTGATCTTTTACAGCACCATTCTTATACAGAAACATATCATAC 2068

Qy 1501 tatgac----- 1506

Db 2069 TATGATGATGAACAGTGTCTAGTTTGAGATCATACAAATCTAAATTCGTGGTGAT 2128

Qy 1507 -----acgggtatgtaccctcccaatat 1531

Db 2129 AAACATCTGTGATTCCTTCTTCTGTGCAGAAAGTACGTGATTTGATTCCTCCAAAT 2188

Qy 1532 gtgtgtgtatgaccttcaacttgaatgatgaaagagtagtaagaagcctatctcatg 1591

Db 2189 GTGTTTGTATGACGTTTACACTTGATGTAAGAGAGTAGTACGAAGCCTATCTCATG 2248

Qy 1592 gattctgacaatcttctccagatctccagatcagaagaagttcgtatgatgatgacctcag 1651

Db 2249 GATTCGTGACAAATCTTCCAGATCCAGCATCGAAGTTCTAAATGGATGATGACCTTCAG 2308

Qy 1652 aaaagtgtgataatgataatgtagtctgtctccaggaataaataatgatgacatat 1711

Db 2309 AAAGATATGATATATATATGATATGATATGTTTCTCTCCAGCAATTAATAATGATGATAT 2368

Qy 1712 ccgtgtgttgaaatgcttcaatcaatgatacaatgltcacaanaatgagacagataatcaat 1771

Db 2369 CCGTGTGTGATGCTTATCAATGATCATGATGATGATGATGATGATGATGATGATGATGAT 2428

Qy 1772 tgcataagattttgacacacagttgcaagaagtagtaactaa 1816

Db 2429 TCCATACAGATTTTGTACACCAACAGTGTGCAAGATGTAATCTAA 2473

RESULT 6

AK026234

LOCUS AK026234 1752 bp mRNA PRI 29-SEP-2000

DEFINITION Homo sapiens cDNA: FLJ22581 fis, clone HS102573.

ACCESSION AK026234

VERSION AK026234.1 GI:10439019

KEYWORDS oligo capping; fis (full insert sequence).

SOURCE Homo sapiens human small intestine cDNA to mRNA, clone Jib:HS1 clone:HS102573.

ORGANISM Homo sapiens

REFERENCE Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.

AUTHORS Watanabe, K., Kunagai, A., Itakura, S., Yamazaki, M., Tashiro, H., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Nakamura, T., Isogai, T. and Sugano, S.

TITLE NEDO human cDNA sequencing project

JOURNAL NEDO unpublished (2000)

REFERENCE 2 (bases 1 to 1752)

AUTHORS Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T., Shibahara, T., Tanaka, T. and Nakamura, Y.

TITLE Direct Submission

JOURNAL Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)

COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing; Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).

FEATURES

source 1..1752

Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

SOURCE		Homo sapiens teratocarcinoma cell_line:NT2 CDNA to mRNA, clone.lib:NT2RM2 clone:NT2RM2001544.
ORGANISM		Homo sapiens
REFERENCE		Eukaryota; Metazoa; Chordata; Carinata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (sites)
AUTHORS		Isoqal,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Kodatori,A., Sudo,H., Magatsura,M., Hosouri,T., Kaku,Y., Shidara,H., Kondo,H., Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K., Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuhio,Y., Niwomiya,K. and Iwayanagi,I. NEBO human cDNA sequencing project Unpublished (2000) 2 (bases 1 to 2631)
TITLE		Isogal,T. and Otsuki,T. Direct Submission Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases. Takao Isogal, Heli Research Institute, Genomics Laboratory: 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952) NEBO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan: cDNA full insert sequencing. Research Association for Biotechnology: cDNA library construction. 5' - 3' end one pass sequencing and clone selection Heli Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.
COMMENT		Location/Qualifiers 1..2631 /organism="Homo sapiens" /db_xref="taxon:9606" /cell_line="NT2" /cell_type="teratocarcinoma" /clone="NT2RM2001544" /clone.lib="NT2RM2" /note="cloning vector: pME18SFU3; mRNA from uninduced NT2 neuronal precursor cells." 24..1928 /code="unnamed protein product" /odon_start=1 /protein_id="BA9A1568.1" /db_xref="GI:7022355" /translation="MSLVPAATNYLYTLPNLQGGTIYVNVGVKFFRPRLYSKGTDVOC SVYIVDQTNRKLCLFSPGNEYALPIITKNGLDIYRFRHLKIYYKKETOGITSSGF SLIEFGDGAPILIPRISKTIFNFTEDHKWYELRWASTHSPTLLKCDVOPMK YEFDLIDQLAKAEVDGASFLKWDRTPPEPMKVYLQDLVEGDLSHHRLOTLTI DILAYDNCHLVKRAWSFKVGSFLRIYLSHTLKLOSMSSENOTMLLEFHNGSYCGJBI VLPSNSVDYDLDKRDESNANTANOHSUYICOSEPDSDFPGSGSYSLVEKPCHEOEV TILDHQYLERLTICALIKOKAPOOYIRIALRSYKPRFLFOVSXKLBKCHLOEVAL HEGDLDIIFODGATRKPYVKRONTSLYKSIIMTKIKGRVAHVAFVNNOGNIPLPSN CLLIIBGSTLSEIKLSKRNFSVIPRGSHDELIDLSAPFLIGTITHHYCKRSS LRSTIONSLVDKTSWIPSSVAELGIVLPDVFPMTITLDDGLEGVAIIMDSKRF QIPASEVLMDDIDLQKSDVMIDMFCPPKIKIDAVPMLECFIKSYVTNGINDIOCYOI FTIPAEVDVI"
CDS		
BASE COUNT	826 a 473 c 491 g 841 t	
ORIGIN		
Query Match	94.4%:	Score 1713.8; DB 9; Length 2631;
Best Local Similarity	95.2%:	Pred. No. 0;
Matches 1814:	Conservative 0; Mismatches 2; Indels 89; Gaps 1	
D0	1 atgtcttgattccagcaacaatatatatataaccctgaataactaaagtgtg 60 	
D1	24 ATGTCTTGATTGCCACCAATATATATATACACCCTGAATCAACTTAAGGTTGT 83 	
D2	61 acaattgcacatgatcattgttgytaagttcctaagaagcccccatatctaagcaaaa 120 	
D3	84 ACAATTGCACATGTCATGCTGTGTCGAAGTCTTTAAAGCCCCCATATCTAAGCAAAA 143 	
D4	121 actgattatcttcagtttgttaactatgttggaccagacaagaatgaactaactgttcctg 180 	

Db	144	ACGATATATGCTCAGTGTGAACTATGTGGACCGACAAATGTAAACTAACTAATTGGCTG	203
Oy	181	ctctttagtggaaacctatgaagcccttcctaatttataaaatvggagatatgttcgc	240
Db	204	CTCTTTAGTGGAAACATGTAAGCCCTCCAAATAATTTATAAAATGGAGATATGTTCGC	263
Oy	241	tttcaagagctgaagaattcaaglatataaaagagactcaggtatataccaagcttcggc	300
Db	264	TTTCACAGCGGTGAAGTTCAAGATATATAAAGAGAGACTACGGGTATACCAGCTTGGC	323
Oy	301	tttgcattcttgaagctttgaaggaaactttggagccctatacactcgcacttcaagc	360
Db	324	TTTGCAATCTTTGACGTTTGAGGAACTTTTGGAGGCCCTATCATACCTCGACATTCAGC	383
Oy	361	aagatatttaactcaactactagagaccaaaatgtagaagccttagctgtttggaca	420
Db	384	AAGATATTTTAACTTCACTACTAGAGACCACAAATGTATGAAGCTTAGCTTTGGGCA	443
Oy	421	tctactcaatgltcaacgctcttggacattactaataattgltgltatgltcaagcaatgag	480
Db	444	TCTACTATATGTACACCGTCTTGAGACATTTAAATTTGTGATGTTCACCCATGGAG	503
Oy	481	tatttgaacctgtcgtcagcctcttggccaagagaatgtagcggagactatcttctt	540
Db	504	TATTTTAACTGACTGTCTAGCTCTTGGCCAAAGCAAGATGGACGAGCATATTCTT	563
Oy	541	ctaaagatgtagatgagcaccagacacattccactcttggagsgcttaataacaaac	600
Db	564	CTAAAGGTATGGATGGCACACAGACACCATTTCCTTGGAGAGTCTTATATACAAAC	623
Oy	601	ctgtctcttgaaggaatttaagtaacatccatccgcctcaaaactctacaatagacatt	660
Db	624	CTGTCTTTTAAAGGTATTTAACTCAATCCATCGGCTACAAATCTGACATAGACATT	683
Oy	661	ttagltcagcaataacatgltcaltgttggcaagaatctcgaaggttgaagcttctttaga	720
Db	684	TTAGTTTACATTAACCATGTTCATGTGGCAAGATCTCTGAAGTTGGAACTTCTTGA	743
Oy	721	atctatagcctctatccaacttcaatlaagaattcgaagaatcgaacaagtgttaagt	780
Db	744	ATCTATAGCCTTATACCAAACTTCATATAATGAATTCAGAAATTCACAACTGTTAAGT	803
Oy	781	tttagagttcatctctatgtaggttaccagttacggtcggagatcagggtcttgcgaag	840
Db	804	TTAGAGTTTCATCTTCATGAGAGTACCATTTACGGTGGGGAATCAGGGTCTTGGCAAA	863
Oy	841	agtaactcgtatgtagtcaactctgaaaaggaattgaatctgcgaattgacagccaat	900
Db	864	AGTAACCTCATGTGGATCACTCACTGAAGAAAGATTTTGAATTCGAAATTTGACAGCCAT	923
Oy	901	cagatctcagatgttactcgttcaataagaacctgacgaagcttccaaagcttcgtagtca	960
Db	924	CAGCATTCAGATGTATCTGTCAATAGAACTCGACGACAGCTTTCGAAAGCTGTGGATCA	983
Oy	961	gtatcattatacgaagtagaagaatgltcaacagctactcgttacaatacttacaagatc	1020
Db	984	GTATCATATATACGCGTGAAGAAATGTCAACAGCTATCTGTCAATATCTTACACATAT	1043
Oy	1021	cagatattgagaggaaccaactatgccaatttgaacaaaagacctccaacaataac	1080
Db	1044	CAGATTTTGGAGGACACCATATGTGGCCATTTTAAACAAAAGCTCTTCAACAAATAC	1103
Oy	1081	cgatccgagcaaaattgagggtacataaagcccaagaagactattcagtcgtttaaactt	1140
Db	1104	CGCATCGAGCAAAATTTGAGTCAATATTAAGCCAGAAAGCTATTTCAGTCTTTAACTT	1163
Oy	1141	cattgcccataatgltcatgtgtcgcaagaagtccacatgagggcggaatttgataataat	1200
Db	1164	CATGTCCCTAATATCTATTTGTCTGCAAGAAAGTTCACATGAGGGCCATTTGGATTAATT	1223
Oy	1201	tttaagatggttgaactaaaccagagatgctaaacttaaaaataacatcatatgat	1260

```

Db 731 CTTGTTCTTGAAGGTGATTTAAGTCACATTCACGCTCAAAATCTGACAAATAGACATT 790
Oy 661 tttagtctagaataaccatttctatctgtgcaagatctctgaaggttggaagcttctctaga 720
Db 791 TTAGTCTAGCAATACCATGATCTGATGCGAAGATCTCTGAAGCTTGGAAGCTTTCTTAA 850
Oy 721 atctatagcttcatatccaaacttcaatcaatgaatcagaagaatcagatgaat 780
Db 851 ATCTATAGCTTCATACCAAACTTCATCAATGATGAGATGAGCAATGACAAATGTTAA 910
Oy 781 tttagatctcatctctctatctgaggttaccagttacagtcgaggaatcagaggtcttccagaa 840
Db 911 TTAGAGCTTTCATCTTCATGAGGTACCATGAGTTCAGGTCGGGAATCAGGCTTCCAGAA 970
Oy 841 agtaactctgaatgtgatacactgaagaaagattagaatctgcgaatttgaaagccat 900
Db 971 AGTAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1030
Oy 901 cagcatctcagatgttctcgtcactcagaactcagacagcttctcgaagctctgagatca 960
Db 1031 CAGCATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1090
Oy 961 gatactatctagaggttgaagaagatgtcagagctctcgtcactcacttaccagatcat 1020
Db 1091 GTATCATTTATACGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1150
Oy 1021 cagatcttgaagaagacacactatgtccacttctgaagaagaagctctcgaagatcat 1080
Db 1151 CAGTATTTGAGAGGAGACACCATGATGATGATGATGATGATGATGATGATGATGATGAT 1210
Oy 1081 cagcatctcagaagaagatgtgagatctatgaagccagacatctcagtcgttgaact 1140
Db 1211 CCATCTCCGAGCAAAATGTGAGGTGATGATGATGATGATGATGATGATGATGATGAT 1270
Oy 1141 cagtcgccaagtgtcacttctcgaagaagttccagagagagagcttctgagatcat 1200
Db 1271 CATTCGCCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1330
Oy 1201 ttccagatctgtgcaactaaacccagatgtcgaagctcacaatcactcactatgat 1260
Db 1331 TTTCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1390
Oy 1261 tcaaaactctgacacacacacacacacacacacacacacacacacacacacacacacac 1320
Db 1391 TCAAAATCTGACCCCTAAATCAAAATCAAAATCAAAATCAAAATCAAAATCAAAATCA 1450
Oy 1321 aataatgtatctctccgcttctcaaatgaatgtcacttctgataagaagagatcact 1380
Db 1451 AATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1510
Oy 1381 agtgaatctgcaaacctctcgaagaagtttaagtgttaattctcgtgagatctgagc 1440
Db 1511 ACTGAAATTTGCAAACTCTGCAAAAGTTAATGATGATGATGATGATGATGATGATGAT 1570
Oy 1441 gaagacactggaacttctgacactcagcaccacttctatataagaagaacatacag 1500
Db 1571 GAAGACCTGGAACCTTTTGACACTTTTGACACTTTTGACACTTTTGACACTTTTGACAC 1630
Oy 1501 tatg----- 1504
Db 1631 TATGATGATTAACAGTGTCTAGTTGATGATCATTCAAAATCTAATTCCTGTTGAT 1690
Oy 1505 -----gagctgagatctgtacccctcaatat 1531
Db 1691 AAAACATCGTGAATTCCTTCTCTGTGCGAGAAAGCTGGGATGATGATGATGATGATGAT 1750
Oy 1532 ggtctgtctatgacacttctcactctgatacgaacagagatctactagaagccatctcag 1591
Db 1751 ggtttgttatgacacttctcactctgatacgaacagagatctactagaagccatctcag 1810
Oy 1592 gattctgaanaatctctcagatctcagacatcagaagctctcagatgagatcactcag 1651

```

```

Db 1811 GATTCGCAAAATCTCTCCAGATTCACGATCAGAAAGTTCGATGATGATGATGATGATGAT 1870
Oy 1652 aaaagtgtgatacagatcagatgatactgttctcctcgaagaataaaatgatacagat 1711
Db 1871 AAAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1930
Oy 1712 ccgtgtggaatgtctcactcagatcagatcagatcagatcagatcagatcagatcagat 1771
Db 1931 CCGTGTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1990
Oy 1772 tgcatacagatcttgcacacacacagttcagaagaatgatacctaa 1816
Db 1991 TCGTATCAGATTTTTCACACACAGTTCAGAAAGTGTGATGATGATGATGATGATGATGAT 2035

```

RESULT 3
AK022580
LOCUS AK022580 2383 bp mRNA PRI 29-SEP-2000
DEFINITION Homo sapiens CDNA FLJ12518 f1s, clone NT2RM2001805.
ACCESSION AK022580
VERSION AK022580.1 GI:10434050
KEYWORDS oligo capping; f1s (full insert sequence).
SOURCE Homo sapiens testicular carcinoma cell_line:NT2 CDNA to mRNA,
clone_11b:NT2RM2 clone:NT2RM2001805.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y.,
Matsunawa,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,
Makamatsu,A., Nakamura,Y., Nagahashi,K., Masuho,Y. and Sasaki,N.
NEO human cDNA sequencing project
Unpublished (2000)
2 (bases 1 to 2383)
Isogai,T. and Otsuki,T.
TITLE Direct Submission
JOURNAL Submitted (23-AUG-2000) to the DDBJ/EMBL/GenBank databases. Takao
Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT
NEO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction; 5'- & 3'-end one pass sequencing and clone selection:
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
source
Location/Qualifiers
1. 2383
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="NT2"
/clone="NT2RM2001805"
/note="Cloning vector: pME18SFL3-mRNA from uninduced NT2
neuronal precursor cells."
439..2343
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB14110.1"
/db_xref="GI:10434051"
/translation="MSLVPRNTNYITPLNDLKGCTIVNYGVKFFRPPPLSGTDYIC
SVTVIVDQTNVLCLEFSNGVEALPIYKNGDIVRFRRLKIOYKKETOGTISGGA
SLTEGLIGAPIIPRTSSKYNFTTEDEHKVEALRWASTHMSPTLAKCDVOPMO
YFDLTCOLGKAEVDGASFLKVDGRTPEPSPRWVLIODLVLEGLDSIHRQNLTI
DIIVYDNHVAHARSILKGSFLRTSLHKLQSMSENOTMISLEFLHGTSTGGRIR
VLSPNSNDVDLKDIDESANLTAOHSDVICOSSPDSFSSGSVSLYEVRCOOSA
TILTDHOILENTPLCALIKQAPQOYRIKAKLSYKRRRLFOYKRLKCPCHLOEYR
HEGDUDITFDGATKTDVKQNTSLDSKTIWTKNOKRVAHVAVHFNKNGILPLSNE
CLLIIEGTTSEIKLSNKFNSVPIPVSGHEDELLDLSAPFLQGTIHYGCKOCS
LRSTIONSLVDKTSWIPSSVAEALGIVPLOYVYVMTFTLDGTGVLEAYLMDSKFF

Email: c9apbs-r@mail.nih.gov
 Tissue Procurement: ATCC/DCTD/DTF
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc-mcgenhgrl.nih.gov
 Shenchenko, Y., Weltherby, K.D., Beckstrom-Sternberg, S.M.,
 Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
 Lim, M., Maduro, O.L., Mastello, C., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Snyder, B., Stanlidrop, S., Thomas, P.J.,
 Tjongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,
 Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 15 Row: 9 Column: 18
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 10434050.

FEATURES

Source

Location/Qualifiers

1..2029

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="MGC:10280 IMAGE:3955573"

/tissue_type="Skin, melanotic melanoma."

/clone_lib="NIN_MGC_20"

/lab_host="DH10B-R"

/note="Vector: pOTB7"

67..1971

/codon_start=1

/product="Unknown (protein for MGC:10280)"

/protein_id="AAH02923.1"

/db_xref="GI:12804139"

/translation="MSLVPAATNYYITPLNOLKGGTIIVVGVYKFFKPPYLSKGTDC

SVVITVDQTNWKLCLFSGNTEALPIIYKNGDIYRFLKLOVKKETOGITSSGFA

SLFEGTIGAPRIPTSSKYFNFTEDHKWALNVAHSTHSPMTLLKLDVOPMO

YEDLCQLGAEVGVASFLKLVMDGTRTPSPSVLVLDVLEEDLSHILHOLTI

DLVYDNHVAHRSKLSGFLIYSLHTRLOSNGENOTMLSEFLHGGISYSGIR

VLPESSNDVOLDKLESNANLTANOSHVOICSEPDSPSPSGSVSLYEVECCOOLSA

TILDYOLERPLCAITIKOKAPOOYRIKATRSKPRRLTOSVYLRHKCLUDEVP

HEGDDIIFQBGATITPPYKQNTISLYRSKIPITIKNOGRKVAHFNKNGILPLNSE

CLLIEGGILSEICKLSNKFNSVIVPRSGHEDLELDLSAPFLIGTIIHHGCKQCSS

LRSIONLSLVDKTSMIPSSVAEALGIPLVQVFMPTFLDVGTVLVAIYLMDSKRF

QIPASEVLMDDLDKSVDMIMDMCEPCGICIKIDAYPMLECFIKSYNVNTGNDOLCYOI

FDTYAEVYI"

BASE COUNT 655 a 391 c 386 g 597 t
 ORIGIN

Query Match

94.5%; Score 1717; DB 9; Length 2029;

Best Local Similarity 95.3%; Pred. No. 0;

Matches 1816; Conservative 0; Mismatches 0; Indels 89; Gaps 1;

0Y 1 atgtcttgggtccagcaacaattatataatacaacccctgaactaactgaagtgat 60
 |||||
 Db 67 atgtcttgggtccagcaacaattatataatacaacccctgaactaactgaagtgat 126
 |||||
 0Y 61 acaatgtcaatgtctatgtgtgtgaagttcttaagcccccatactaaagaaga 120
 |||||
 Db 127 ACAATTGTCATAGTCTATAGTGTGTGAAGTTCTTTAAGCCCCCATATATCAGCAAGA 186
 |||||
 0Y 121 actgattatgtcagttgtaactatgtgacacagaacaatgtaaacacttgctg 180
 |||||
 Db 187 ACTGATTATGTGCTACGTGTAATGTGTGACACAGCAAAATGTAAACTACTTCC 246
 |||||
 0Y 181 cctcttaagtggaactaagaagcccttccaataattataaaaaatggaalatgtgcg 240
 |||||
 Db 247 CTCTTTAGTGAAGAACTATAGAGCCCTTCCAAATATATTATAAATGGAATATGTGTTCC 306
 |||||
 0Y 241 ttccacagctgaagatccaagataataaagaagagcctcaggtatccaccagctctgc 300

Db 307 ttccacagctgaagatccaagataataaagaagagcctcaggtatccaccagctctgc 366
 |||||
 0Y 301 ttgcacattgaagcgtttgaaggaacttggagccctataactcgcactgaagc 360
 |||||
 Db 367 ttgcacattgaagcgtttgaaggaacttggagccctataactcgcactgaagc 426
 |||||
 0Y 361 aagtaatttaactcaactcgaagcacaacaaatggtgaagcctcgtttgggca 420
 |||||
 Db 427 AAGTATTTTAACTTCTACTAGACACCAAAATGTAAGACCTTACGTTGGGCA 486
 |||||
 0Y 421 tctactcataatgacccgctctgagcattactaaatgtgtgatlgaatcgaatgag 480
 |||||
 Db 487 TCTACTCATATGTACACCGCTTGACCATTAATAATTGTGATGTTCAGCAATGAGC 546
 |||||
 0Y 481 tatltagcactgaactgtgaagccttgggcaagaagaaggtgagagcattcttc 540
 |||||
 Db 547 TATTTTGACCTGACTTGTGAGCTTTGGGCAAGAGAGAGAGGAGCATATTTCTT 606
 |||||
 0Y 541 ctaaagataatgagagcagagacacacattccactctggaaggtcttaataagac 600
 |||||
 Db 607 CTAAGAGTATGGATGGACACAGACACCATTTCCATCTTGAGAGTCTTAATACAGAC 666
 |||||
 0Y 601 ctgttcttgaaggtgatttaagtcacalcacggctacacaaatctgacaatagacat 660
 |||||
 Db 667 CTGTGTTCTGAAGGTGATTTAAGTACATTCATCGGCTACAAATCTGCATATAGCAT 726
 |||||
 0Y 661 ttagtctgaataaaccatgttcatgtgtggaagatctctgaaggttggaaagctttctga 720
 |||||
 Db 727 TTAGTCTACGATTAACCATTTGATGTGGCAAGATCTCTGAGGTGGAAGCTTTTGA 786
 |||||
 0Y 721 atctatagccttcatcacaacttcaatcaatgaatcagaagaatcaagaatgttaag 780
 |||||
 Db 787 ATCTATAGCTTCATATACCAACTTCAATCAATGAANTCTGAGAAATCAGACAAATGTAAGT 846
 |||||
 0Y 781 ttagagttcatcttcaatgagagtaacagctacaggtctggggaatcaagggtcttgcagaa 840
 |||||
 Db 847 TTAGAGTTTTCATCTTATGAGGATGACCATTCGCGGGGAGATAGGCTTTGGCAAGA 906
 |||||
 0Y 841 agtaactcgaatgtgataactgtaaaaaagattagaactctgcaaatcttgaacagcaat 900
 |||||
 Db 907 ACTTACTTGATGTGATCACTGAAAAAGATTTAGAACTCTGCAAATTTGACAGCCAT 966
 |||||
 0Y 901 cagcaatcagaatgtatctgtcaatcagaacctgacagacagcttccaagctctgagaca 960
 |||||
 Db 967 CAGCAATTCAGATGATATCTGTCATACAGAACCTGACAGAGCTTTCCAAAGCTTGATGATCA 1026
 |||||
 0Y 961 gtaactatatacgaaggaagaatgtaacagagctatctgtctataacttaacagatcat 1020
 |||||
 Db 1027 GTATCATTAATACGAGGTAGAAAAGATGTCAACAGCTATCTGTACAAATCTTAACAGATCAT 1086
 |||||
 0Y 1021 cagatattggaagagacacacactatgtgcatattggaacaaagaactctccaacatlac 1080
 |||||
 Db 1087 CAGTATTTGAGAGGACACACACTATGTGTCATTTTGAACCAAAAGCTCTCAACAAATATAC 1146
 |||||
 0Y 1081 cgcattccgaagcaaatgaggttcataaagcccgagaagacttccaagtcgttaactt 1140
 |||||
 Db 1147 CCGATTCGAGCAAAATGTAGGTATATAAGCCCAAGACTATTTTCACGTGTGTAACCTT 1206
 |||||
 0Y 1141 catgcccctaattgcatattgtctgcaagaagttccaatgagagcgtattgataaatt 1200
 |||||
 Db 1207 CATGCGCCTAATATGTCATTTGCTGCAAGAAGTTCCATGAGAGGCGGATTTGGATTAAT 1266
 |||||
 0Y 1201 ttccaagatgtgcaactaaacccccaatgtccaagtccaataatatacatatgat 1260
 |||||
 Db 1267 TTTAGAGATGTGCAACTTAACCCAGATGTCAAGCTCAAAATATCATTTATATATAT 1326
 |||||
 0Y 1261 tcaaaatctggaacacataaaatcaaaagagcaaaagtagaggtcatttttggaa 1320
 |||||
 Db 1327 TCAAAAATCTGACCACTAAATATCAAAAAGAGCAAAAAGTGAAGTTCATTTTGTGAAA 1386
 |||||
 0Y 1321 aataatgtaattctccgcttccaatgtaattctactttgataagaagagtaactc 1380
 |||||

Db 663 www..... 722
Oy 1722 aatgctcatcaagtcatacaatgtaacagataatcgaatgctacaga 1781
Db 723 www..... 782
Oy 1782 ttcttgaccacac 1794
Db 783 wwwgcccccc 795

RESULT 15

AAFS8257

ID AAF58257 standard; DNA: 936 BP.

XX

XX

XX 24-APR-2001 (first entry)

XX Oligonucleotide D1954.

XX Electron-transfer group; ETM; mismatch; genotyping;

KM gene expression; ss.

XX

XX Synthetic.

XX MO200107665-A2.

XX

XX 01-FEB-2001.

XX 26-JUL-2000; 2000MO-US20476.

XX 26-JUL-1999; 9905-0145695.

XX 17-MAR-2000; 2000US-0190259.

XX (CLIN-) CLINICAL MICRO SENSORS INC.

XX

XX Umek RM;

XX

XX WPI; 2001-159728/16.

XX

XX Nucleic acids containing electron-transfer group, useful as labels in

XX PT hybridization assays, e.g. for genotyping, allowing repeat analyses on

XX PT a single surface

XX

XX Example 6; Page 127; 159pp; English.

XX

XX The present invention relates to a composition comprising two nucleic

XX CC acids each containing an electron-transfer group (ETM) having

XX CC different redox potentials. The invention is used for electronic

XX CC detection of nucleic acids, especially of substitutions (mismatches)

XX CC and single-nucleotide polymorphisms, e.g. for genotyping,

XX CC monitoring gene expression.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Oy 1182 gggcgtttgagataatlttccagatggtgcactaaacccagatgtaagctaca 1241
Db 183 www..... 242
Oy 1242 aatatacatcttataatgatacctggaacactaaatcaaaagcgcaagat 1301
Db 243 www..... 302
Oy 1302 agcagttcatttctgaaataatgtaatcccgcttcaatgtaagctacttt 1361
Db 303 www..... 362
Oy 1362 gatagaagaggtacactcagttgaattgcaactctcgaaagtttaagtgta 1421
Db 363 www..... 422
Oy 1422 tctgtgagatctgcccgaagacgtggaactttggaacttccagacacttctat 1481
Db 423 www..... 482
Oy 1482 acaaggaacatacatcactatgacgtggtatgtacccctcaatagtgtgta 1541
Db 483 www..... 542
Oy 1542 tgaccttactatgatagtaacggagtaactagaagcctatccatgattctgaca 1601
Db 543 www..... 602
Oy 1602 aattctccagattccagatcagaagttctgatagtgacacttcagaagtglog 1661
Db 603 www..... 662
Oy 1662 atatgatacgtatgatttctccccaagaataaattgatacatccgtgtg 1721
Db 663 www..... 722
Oy 1722 aatgctcatcaagtcatacaatgtaacagataatcgaatgctacaga 1781
Db 723 www..... 782
Oy 1782 ttcttgaccacac 1794
Db 783 wwwgcccccc 795

Search completed: April 9, 2002, 22:18:59
Job time: 7678 sec

Query Match 11.1%; Score 201; DB 22; Length 675;
Best Local Similarity 100.0%; Pred. No. 1.1e-45;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1616 ccagatcagaagttctatgatgatgtactctcagaataatgtgagatatacatgat 1675
Db 1 ccagatcagaagttctatgatgatgtactctcagaataatgtgagatatacatgat 60
Oy 1676 atgtttgtccctcagagaataaaltgatcatatccgtgtgtggaatgtcctcaag 1735
Db 61 atgtttgtccctcagagaataaaltgatcatatccgtgtgtggaatgtcctcaag 120
Oy 1736 tcatacaatgtccacaatagagacagataatcaattgtcatcagattttgacaccaca 1795
Db 121 tcatacaatgtccacaatagagacagataatcaattgtcatcagattttgacaccaca 180
Oy 1796 gttgcagaagatgttaactaa 1816
Db 181 gttgcagaagatgttaactaa 201

RESULT 12

AA139284
ID AA139284 standard; DNA; 475 BP.

AC AA139284;

DT 17-OCT-2001 (first entry)

DE Probe #7970 used to measure gene expression in human placenta sample.

KW Probe: microarray; human; placenta; antenatal diagnosis;
genetic disorder; ss.

OS Homo sapiens.

PN W0200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632386.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488897/53.

PT Human genome-derived single exon nucleic acid probes useful for

PS analyzing gene expression in human placenta -

PS Claim 25; SEQ ID No 7970; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENP).

CC The present sequence is one such probe. The probes are useful for

CC producing a microarray for predicting, measuring and displaying gene

CC expression in samples derived from human placenta. The probes are useful

CC for antenatal diagnosis of human genetic disorders.

XX Sequence 475 BP; 191 A; 70 C; 69 G; 145 T; 0 other;

SO Query Match 8.1%; Score 146.8; DB 22; Length 475;

Best Local Similarity 92.8%; Pred. No. 9.1e-31;
Matches 154; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Oy 1140 tcatgtccccaatagtcatttctgtcagaagaagttccacatgagggatttgataat 1199
Db 310 tcatgtccccaatagtcatttctgtcagaagaagttccacatgagggatttgataat 369
Oy 1200 ttctcagaatgtgtcacaactaaacccagatgtcacaagtacaaatatacatataga 1259
Db 370 ttctcagaatgtgtcacaactaaacccagatgtcacaagtacaaatatacatataga 429
Oy 1260 ttcaaaaatcttgagccactaaatacaaaaaggacgaagaataga 1305
Db 430 ttcaaaaatcttgagccactaaatacaaaaaggacgaagaataga 475

RESULT 13

AAF58252
ID AAF58252 standard; DNA; 936 BP.

AC AAF58252;

DT 24-APR-2001 (first entry)

DE Oligonucleotide D1835.

KW Electron transfer group; ETM; mismatch; genotyping;

gene expression; ss.

OS Synthetic.

PN W0200107665-A2.

PD 01-FEB-2001.

PF 26-JUL-2000; 2000WO-US20476.

PR 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.

PA (CLIN-) CLINICAL MICRO SENSORS INC.

PI Umek RM;

DR WPI; 2001-159728/16.

PT Nucleic acids containing electron-transfer group, useful as labels in

PS hybridization assays, e.g. for genotyping, allowing repeat analyses on

PS a single surface -

PS Example 6; Page 127; 159pp; English.

XX The present invention relates to a composition comprising two nucleic

CC acids each containing an electron-transfer group (ETM) having

CC different redox potentials. The invention is used for electronic

CC detection of nucleic acids, especially of substitutions (mismatches)

CC and single-nucleotide polymorphisms, e.g. for genotyping,

CC monitoring gene expression.

XX Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

SO Query Match 6.7%; Score 121; DB 22; Length 936;

Best Local Similarity 1.3%; Pred. No. 1.7e-23;
Matches 10; Conservative 484; Mismatches 299; Indels 0; Gaps 0;

Oy 1002 tacaatactacagatcagatatttgagagagacacactatgtgccatttgaaca 1061
Db 3 tacaatactacagatcagatatttgagagagacacactatgtgccatttgaaca 62

Oy 1062 aaaagtcctcacaacatcacgcattccgagcaaatgtgagtcataaagccagaagact 1121
Db 63 aaaagtcctcacaacatcacgcattccgagcaaatgtgagtcataaagccagaagact 122

OY 1505 -----gcactgggtatgtacccctccaatagtgtttgtatgacctta 1550
CC |
CC |
CC |
DB 601 ctctctgagcagaagcagctgggtattgtaccctccaatagtgtttgtatgacctta 660
OY 1551 cacttgatattggaacagagatcagaagcctatctctgtatcttgcacaattcttc 1610
CC |
CC |
CC |
DB 661 cctctgagatgtggaagaaggaactagaagcctatctctgtatcttgcacaattcttc 720
OY 1611 agattccagcatcagaagctctgagtgtatgagatgagacccctcagaagaagtgtgatatac 1670
CC |
CC |
CC |
DB 721 agattccagcatcagaagctctgagtgtatgagatgagacccctcagaagaagtgtgatatac 780
OY 1671 tggatattgtttgtctctcaggaataaaatgtgatacatccgtgtgttggaatcttca 1730
CC |
CC |
CC |
DB 781 tggatattgtttgtctctcaggaataaaatgtgatacatccgtgtgttggaatcttca 840
OY 1731 tcaagtcatacatgtcacaatgtgacagatataatcttgcatttcgattttgaca 1790
CC |
CC |
CC |
DB 841 tcaagtcatacatgtcacaatgtgacagatataatcttgcatttcgattttgaca 900
OY 1791 ccacagctgcagaagaatgtatctaa 1816
CC |
CC |
CC |
DB 901 ccacagctgcagaagaatgtatctaa 926

RESULT 6

AAH05303
ID AAH05303 standard; cDNA; 576 BP.

AAH05303;

26-JUN-2001 (first entry)

Human cDNA clone (5'-primer) SEQ ID NO:2138.

Human; primer: detection: diagnosis; antisense therapy: gene therapy; ss.

Homo sapiens.

EP:074617-A2.

07-FEB-2001.

28-JUL-2000: 2000EP-0116126.

29-JUL-1999: 99JP-0248036.

27-AUG-1999: 99JP-0300253.

11-JAN-2000: 2000JP-0118776.

02-MAY-2000: 2000JP-0183767.

09-JUN-2000: 2000JP-0241899.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

WPI: 2001-318749/34.

Claim 1: SEQ ID 2138; 2537bp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end

sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 576 BP: 167 A; 113 C; 114 G; 173 T; 9 other;

Query Match 30.0%; Score 545; DB 22; Length 576;
Best Local Similarity 98.7%; Pred. No. 8, 2e-141;
Matches 545; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 atgtcttggttccagcaacaattatataataacacccctgaacttaaggtgtg 60
DB |
DB 25 atgtcttggttccagcaacaattatataataacacccctgaacttaaggtgtg 84
OY 61 acaatgtcaatgtctatgtgtgtgtgaagttctttaaagccccaatcagaaga 120
DB |
DB 85 acaatgtcaatgtctatgtgtgtgtgaagttctttaaagccccaatcagaaga 144
OY 121 actgattattgtcctgaattgttaactatgtgagcagaacaatgttaactaactgtcgt 180
DB |
DB 145 actgattattgtcctgaattgttaactatgtgagcagaacaatgttaactaactgtcgt 204
OY 181 cctcttagtggaaacatgaagccttccataattataaaatgagatattgtcgc 240
DB |
DB 205 cctcttagtggaaacatgaagccttccataattataaaatgagatattgtcgc 264
OY 241 ttccacagctggaagatccaagataaaagaagagactcaaggtatccagctctgcg 300
DB |
DB 265 ttccacagctggaagatccaagataaaagaagagactcaaggtatccagctctgcg 324
OY 301 ttgcacatttgcagcttgaaggaacttggagccctcatatcaactctgcaactcaagc 360
DB |
DB 325 ttgcacatttgcagcttgaaggaacttggagccctcatatcaactctgcaactcaagc 384
OY 361 aagattttaactcactaactgagaccacaatgtgaaagccttaactgtttgggca 420
DB |
DB 385 aagattttaactcactaactgagaccacaatgtgaaagccttaactgtttgggca 444
OY 421 tctactcattatgtcaccgctctggacattactaataattgtgagtttaagccaatgcag 480
DB |
DB 445 tctactcattatgtcaccgctctggacattactaataattgtgagtttaagccaatgcag 504
OY 481 tatttgacctgtgtcagcctctgggcaagaagagagagcagcatcatttctt 540
DB |
DB 505 tatttgacctgtgtcagcctctgggcaagaagagagagcagcatcatttctt 564
OY 541 ctaaaagtgatgg 552
DB |
DB 565 ctaaaagtgatgg 576

RESULT 7

AAH07649
ID AAH07649 standard; cDNA; 854 BP.

AAH07649;

26-JUN-2001 (first entry)

Human cDNA clone (5'-primer) SEQ ID NO:4484.

XX 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI: 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -

PS Claim 8: SEQ ID 11187; 2537pp + CD ROM: English.

XX The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dt primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB92893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX

Sequence 2631 BP; 826 A; 473 C; 491 G; 841 T; 0 other;

Query Match 94.4%; Score 1713.8; DB 22; Length 2631;
 Best Local Similarity 95.2%; Pred. No. 0;
 Matches 1814; Conservative 0; Mismatches 2; Indels 89; Gaps 1;

QY 1 atgctcttggtccagaacaattatataatatacccccgaatcaactaaaggtgt 60
 Db |
 Db 24 atgctcttggtccagaacaattatataatatacccccgaatcaactaaaggtgt 83
 QY 61 acaattgcaatgctcactgtgtgtgtgaagttcctaagcccccatctaaagaaaga 120
 Db |
 Db 84 acaattgcaatgctcactgtgtgtgtgaagttcctaagcccccatctaaagaaaga 143
 QY 121 actgattatgtcagttgaactatctgtggaaccagacaaatgtaaaaacttgctgt 180
 Db |
 Db 144 actgattatgtcagttgaactatctgtggaaccagacaaatgtaaaaacttgctgt 203
 QY 181 ctcttgatggaactatgaagccctcccaataattataaaatgagatattgttcgc 240
 Db |
 Db 2c: ctcttgatggaactatgaagccctcccaataattataaaatgagatattgttcgc 263
 QY 241 ttccacagctgaagatcagaatataataaagaagactcagggatcaccagctcggc 300
 Db |
 Db 264 ttccacagctgaagatcagaatataataaagaagactcagggatcaccagctcggc 323
 QY 301 ttgcacattgacgtttgaggaactttggagccctacatcactcgactccaagc 360
 Db |

Db 324 ttgcacattgacgtttgaggaactttggagccctacatcactcgactccaagc 383
 QY 361 aagcatttcaactcactactgaagaccacaaatggtagaagcctacgtgttgagca 420
 Db |
 Db 384 aagcatttcaactcactactgaagaccacaaatggtagaagcctacgtgttgagca 443
 QY 421 tctactcatatgtcacccgtcttggaacttaactaaatltgtgagtgttcagcaatgca 480
 Db |
 Db 444 tctactcatatgtcacccgtcttggaacttaactaaatltgtgagtgttcagcaatgca 503
 QY 481 tatttgacactgtgttcagcctcttggaagcaaaatgtagcagcagatattctt 540
 Db |
 Db 504 tatttgacactgtgttcagcctcttggaagcaaaatgtagcagcagatattctt 563
 QY 541 cttaaagtatggaatggtgacacagacacacattccactctggaaggtcttaatacaag 600
 Db |
 Db 564 cttaaagtatggaatggtgacacagacacacattccactctggaaggtcttaatacaag 623
 QY 601 ctgtctcttggaaggttgaatgaatcagatcagatcagatcagatcagatcagatc 660
 Db |
 Db 624 ctgtctcttggaaggttgaatgaatcagatcagatcagatcagatcagatcagatc 683
 QY 661 ttaactacagataacacatgttcatgtgcaagatcctggaaggttgaagcttcttaga 720
 Db |
 Db 684 ttaactacagataacacatgttcatgtgcaagatcctggaaggttgaagcttcttaga 743
 QY 721 atctatagccttcataccaactcaatcaatgaatcagagaatcagacaatgttaagt 780
 Db |
 Db 744 atctatagccttcataccaactcaatcaatgaatcagagaatcagacaatgttaagt 803
 QY 781 ttaagttcaactcttcagtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgag 840
 Db |
 Db 804 ttaagttcaactcttcagtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgag 863
 QY 841 agtaactctgagtgtgatacactgtgaaaaagatttagatctgcaaatltgacagccaat 900
 Db |
 Db 864 agtaactctgagtgtgatacactgtgaaaaagatttagatctgcaaatltgacagccaat 923
 QY 901 cagatattcagatgtatctgtcactgcaatcgaacccgtgacagacgcttccaagcttgatca 960
 Db |
 Db 924 cagatattcagatgtatctgtcactgcaatcgaacccgtgacagacgcttccaagcttgatca 983
 QY 961 gtatcattatcagaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgag 1020
 Db |
 Db 984 gtatcattatcagaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgag 1043
 QY 1021 cagatattgagaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggt 1080
 Db |
 Db 1044 cagatattgagaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggt 1103
 QY 1081 cgcattcagagcaaatgtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgag 1140
 Db |
 Db 1104 cgcattcagagcaaatgtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgag 1163
 QY 1141 catgcccctaagtgtcattgtgtgcaagaagttccacatgagggcgtttgagatatt 1200
 Db |
 Db 1164 catgcccctaagtgtcattgtgtgcaagaagttccacatgagggcgtttgagatatt 1223
 QY 1201 ttcaaggaatgtgtgcaactaaaccccgatgtcaggtcacaataatcatcatatgat 1260
 Db |
 Db 1224 ttcaaggaatgtgtgcaactaaaccccgatgtcaggtcacaataatcatcatatgat 1283
 QY 1261 tcaaaaatctggaccacttaaaaatcaaaaagagcagaagaatgagatgtcattgtgaaa 1320
 Db |
 Db 1284 tcaaaaatctggaccacttaaaaatcaaaaagagcagaagaatgagatgtcattgtgaaa 1343
 QY 1321 aataatgtatctcccgcttccaatgaaatgtctactcttgatagaagaaggatcacac 1380
 Db |
 Db 1344 aataatgtatctcccgcttccaatgaaatgtctactcttgatagaagaaggatcacac 1403
 QY 1381 agtgaataatggaacactctggaacaaatgtaataatgtatctctgtgagaatctggccac 1440
 Db |
 Db 1404 agtgaataatggaacactctggaacaaatgtaataatgtatctctgtgagaatctggccac 1463


```

/tissue_type="placenta"
/Note="Vector: PM2185FL3"
BASE COUNT      259 a      164 c      166 g      259 t      6 others
ORIGIN

Query Match      28.8%; Score 522.4; DB 10; Length 854;
Best Local Similarity 99.3%; Pred. No. 4.3e-116;
Matches 545; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

OY 1 atgtcttggttccagcaacaataatataataacacccctgaactaacttaaggtgt 60
    |||
DB 132 ATGCTTTGGTTCAGACACAAATATATATATACACCCCTGATCACTTAAGGGTGT 191
    |||

OY 61 acaattgcaatgtctatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 120
    |||
DB 192 ACAATTGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 251
    |||

OY 121 actgattatgtctatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 180
    |||
DB 252 ACTGATTATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 311
    |||

OY 181 ctcttttgtagaacttgaaagccctcccaataatataataatggagatgtgtgtgc 240
    |||
DB 312 CTCCTTTAGTGAACCTATGAAAGCCCTTCCAAATATTTATATAAATGAGATATGTTGTCGC 371
    |||

OY 241 ttccacagctgaaagattcaatataataaagaagagagagagagagagagagagag 300
    |||
DB 372 TTTCACAGGCTGAAGATTCAGATATATAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 431
    |||

OY 301 ttggacatcttgacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 360
    |||
DB 432 TTGGACATCTTTGACGTTTGAGGAGACCTTTGGAGCCCTTATACCTCCACCTTCAGAC 491
    |||

OY 361 aagatttlaacttcaacttgaagagacacacacacacacacacacacacacacacac 420
    |||
DB 492 AAGTATTATTACTTCACTACTGAGACACCAAAATGTTAGAACCTTTACGTGTTTGGGCA 551
    |||

OY 421 tctactcaatgttgcacgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 479
    |||
DB 552 TCTACTCAATGTGTCACCGCTTTTGACATTTGATTAATTTGTTGATGTTGACCAAAATGCA 611
    |||

OY 480 gtaatttgacactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 538
    |||
DB 612 GTAATTTTGACCTACTTGTGCACTTTGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 671
    |||

OY 539 ttctaaag 547
    |||
DB 672 TTCTAAAG 680
    |||

RESULT 11
AV752694      529 bp      mRNA      EST      19-OCT-2000
LOCUS      AV752694 NPD Homo sapiens cDNA clone NPDAEB10 5', mRNA sequence.
DEFINITION      AV752694
ACCESSION      AV752694
VERSION      AV752694.1 GI:10910542
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 529)
Song,H., Peng,Y., Gu,Y., Yang,Y., Gao,G., Xiao,H., Xu,X., Li,N.,
Qian,B., Liu,F., Ou,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu,S., Gu
,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Ye,M., Zhang,O.,
Han,Z., Chen,Z., Hu,R. and Chen,J.
Homo sapiens NPD library cDNA clones
Unpublished (2000)
Contact: Qinghua Zhang
Shanghai Institute of Endocrinology, Rui-Jin Hospital
197 Rui-Jin II Road, Shanghai 200025, P. R. China
Tel: 86-21-64370045(ex.663332)

Fax: 86-21-64743206
Email: mbs@shims.sh.cn
This clone is available at Shanghai Hematology Institute in
Shanghai.
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong.
Location/Qualifiers
1..529
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NPDAEB10"
/clone_lib="NPD"
/tissue_type="placental"
/dev_stage="Adult"
/lab_mosi="SOLR"
/Note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT      168 a      106 c      104 g      150 t      1 others
ORIGIN

Query Match      26.0%; Score 472.8; DB 10; Length 529;
Best Local Similarity 95.4%; Pred. No. 4.2e-104;
Matches 497; Conservative 0; Mismatches 23; Indels 1; Gaps 1;

OY 774 gtttaagtttagagtttcattcatcttcaatgagagagagagagagagagagagag 833
    |||
DB 1 GTTAATTTAGACTTTTCATCTTCATGAGGTACAGTTCAGGTTCGAGGATCAGGGTCTT 60
    |||

OY 834 gccagaaatlaactctgatagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 893
    |||
DB 61 GCCAGAAATLAACCTCTGATGTGTGATCACTGAAAGAGATTAAGATTCGAAATTTGAC 120
    |||

OY 894 agccaatcagacttgcagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 953
    |||
DB 121 AGCCAATCAGCAATTCAGATGTATCTGTCAATCAGACCTGACGACACCTTTCCAAAGCTC 180
    |||

OY 954 tggatcagatcatcatatacagagtagaagaatgtcaacagctatctgtacaatactac 1013
    |||
DB 181 TGGATCAGTATCATTTATACGAGTAGAAGATGTCAACAGTATCTGTCAATATCTTAC 240
    |||

OY 1014 agatcatcagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1073
    |||
DB 241 AGATCATCAGTATTTGGAGAGACACCACTATGTCCATTTTGAACCAAAAGCTCTCCA 300
    |||

OY 1074 acaatcaccgacccgagcaaatltgagtcataataagccagaagaacattcagctgt 1133
    |||
DB 301 ACAATACCGCATCCGAGCAAAATTTGAGTCAATATTAAGCCCAAGACACTTTCAGTCTGG 360
    |||

OY 1134 taacttcattgcccataatgtcatttgcagagaagttccacatagagggcagattga 1193
    |||
DB 361 TAACCTTCATTTGGCTTAATGTTCATTTGCTTCAGAAAGTT-CACATGAGGGCGNATTTGA 419
    |||

OY 1194 tataattttcagagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1253
    |||
DB 420 TATTAATTTTTCAGGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 479
    |||

OY 1254 atagatcaaaaatctgagacccaataaaaatcaaaaagagac 1294
    |||
DB 480 ATATGATTTTAAAAATCTTGACCACTTAATATCGACCGGGC 520
    |||

RESULT 12
AA351459/C      429 bp      mRNA      EST      21-APR-1997
LOCUS      AA351459 Infant brain Homo sapiens cDNA 5' end, mRNA sequence.
DEFINITION      AA351459
ACCESSION      AA351459
VERSION      AA351459.1 GI:2003779
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

```

/tissue_type="rhadomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: muscle; Vector: pOT7; Site:1: EcoRI;
 Site-2: XhoI; cDNA made by oligo-dT priming,
 directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCAAGGAG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 239 a 154 c 145 g 198 t
 ORIGIN

REFERENCE
 AUTHORS NIH-MGC
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC/DC/DTP
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNU)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILNU at: image.llnl.gov
 Plate: LLCM830, row: e, column: 14
 High quality sequence start: 52
 High quality sequence stop: 746.

FEATURES

Query Match 31.7%; Score 575.6; DB 10; Length 736;
 Best Local Similarity 98.4%; Pred. No. 5.5e-129;
 Matches 634; Conservative 0; Mismatches 4; Indels 6; Gaps 5;

ORIGIN

BASE COUNT 275 a 189 c 189 g 267 t

Query Match 30.2%; Score 548.6; DB 11; Length 920;
 Best Local Similarity 92.0%; Pred. No. 1.9e-122;
 Matches 702; Conservative 0; Mismatches 44; Indels 17; Gaps 11;

ORIGIN

BASE COUNT 513 a 307 c 307 g 395 t

Email: c9apbs-remail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLM41258 row: h column: 18

High quality sequence stop: 705.

FEATURES

source

Location/Qualifiers

1..1159

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_image="5106881"

/clone_lib="NIH_MGC_12"

/tissue_type="cervical carcinoma cell line"

/lab_host="DH10B"

/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.4 kb. Library prepared by Life Technologies."

BASE COUNT 353 a 255 c 240 g 311 t

ORIGIN

Query Match

Best Local Similarity 32.3%; Score 590.2; DB 11; Length 1159;

Matches 684; Conservative 0; Mismatches 33; Indels 14; Gaps 6;

Oy 766 cagcaaatgttaagtttagagtttcattccatcabaagagttaccagttcagtcggagatc 825

Db 2 CAGCAAAATGTTAACTTGAAGTTTCACTTCATCGAGAGTACCACTTACGGTGGGAG 61

Oy 826 aggtcttcagcaagaactcgtatgtagatcaactgaaagaagatttagaatcgca 885

Db 62 AGGTTCTTCCAGAAATTAAGTCTGATGATGATCACTGAAAAAGATTAGATCTGCA 121

Oy 886 aattgacagccaatcagcatcagatgtagatcgtatcgtcaatcagacccagcagctt 945

Db 122 AATTGACAGCAATCAGCATTCAGATGTTATCTGCAATCAGAACCTGACGACGCTT 181

Oy 946 ccaagctcgtatcagatcattatatacagagtgtagaagaagatgtaacagctatctgca 1005

Db 182 CCAAGCTCTGATCAGTATATATACGAGTAAAGATGTCAACGATATCTGCTACA 241

Oy 1006 atactacagatcagcatcttggagagagcaccactatgtagcatttgaacaacaa 1065

Db 242 ATACTTACAGTATCGATTTTGGAGAGGACACCATATGTGCTATTTTGAACAAAA 301

Oy 1066 gctcctcaacaatacgcgcatcgcagcaaaattgaggtcatataagccagagaactatt 1125

Db 302 GCTCTCAACAAATACCCATCCGAGCAAAATGAGTCAATATAAGCCACAGACTATT 361

Oy 1126 cagctcgtttaaactcgtccctaaatgta-cttgcctcaagaagttcacatgaggg 1184

Db 362 CAGCTGTAAACTTCAATGCCCTAAATGTCTATTTTCTGCAAGAGTTCCACATGAGG 421

Oy 1185 cga-ttgagataaattttcagagatgtagcaataaaccacagatgtaagctacaa 1243

Db 432 CGATTITGGATATTAATTTTTCAGATGTGTCACATAAACCCCGATGTCAAGCTACAA 481

Oy 1244 atacatatatatgattcaaaaatctggaaccactaaatc-aaaaagagacgaagaatg 1302

Db 482 ATACATATTATATGATTCAAAAATCTGGACACATAAAAATCGCAAAAGACGAAAAAT 541

Oy 1303 gcaatcatttgg-aaaaaatatgtagttcccgcttcaaatgtagtcttactt 1360

Db 542 GCAATTAATTTCTGACACAAATATGTTATTTCTCCGCTTTTCAATGAATGTCTACTT 601

Oy 1361 tgatagaagagtlacatcagtgaaa-ttgcaaaccttcgaacaagtttaagtgt 1419

Db 602 TCATAGAAGAGTACAGTGAATTTTGCMAACTTCGAAACAGTCTAATAGTGA 661

Oy 1420 attctcgtgaga-----ctgagcagagaagactggaacttttgagcctttagcac 1471

Db 662 AATTCCCTGTGAGAGATCTTGCGACACCAAGAACCTGGAACTTTGGACCTTTCAGCAC 721

Oy 1472 catctcttata 1482

Db 722 CAATTCTCTTA 732

RESULT 5

source

LOCUS

DEFINITION

BE884820 588 bp mRNA 20-OCT-2000

601510547F1 NIH_MGC_71 Homo sapiens CDNA clone IMAGE:3912010 5',

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Human.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 588)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: c9apbs-remail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLM9729 row: n column: 11

High quality sequence start: 3

High quality sequence stop: 585.

Location/Qualifiers

1..588

FEATURES

source

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_image="3912010"

/clone_lib="NIH_MGC_71"

/tissue_type="leiomyosarcoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.1 kb."

BASE COUNT 207 a 116 c 105 g 160 t

ORIGIN

Query Match

Best Local Similarity 32.3%; Score 586; DB 11; Length 588;

Matches 586; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 830 tcttgcagaagaagtaactcgtatgtagatcgaactggaagaagatttagaatctgcaatt 889

Db 2 TCTTGCAGAAAGTAAGTATGATGATGATCACTGAAAAAGATTAGATTCGAAAT 61

Oy 890 tgaacgcaatcagcatcagatgtagatcgtatcgtcaatcaaaccttgaagacgaacttccaa 949

Db 62 TGACAGCCAAATCAGCATTTAGATGTATCTGCAATCAAGACCTGACGACACTTTCCAA 121

Oy 950 gctcgtatcagatcattatatacagagtgagaagaagatgtaacagctatctgtacaatc 1009

Db 122 GCTCTGATATAGTATATATATAGAGAGTAAAGATGTAAACAGCTATCTGCTACAAAT 181

Oy 1010 ttacagatcagatatttggagagagacacactatgtagcatttgaacaacaaagctc 1069

Db 182 TTACAGATCATCATTTTGGAGAGGACACCACTATGTGCTCAATTTGAAACAAAAAGCTC 241

Email: filiang@lifetech.com URL:
http://fulllength.invitrogen.com*
BASE COUNT 290 a 176 c 166 g 256 t 1 others
ORIGIN

Query Match 38.5%: Score 699; DB 10; Length 889;
Best Local Similarity 89.8%: Pred. No. 7.3e-159;
Matches 798; Conservative 1; Mismatches 1; Indels 89; Gaps 1;

```

Oy 871 gattagaatcgcgaattgacagcccaatcagcattcagatggtatctgtaacacaa 930
    |||||
Db 1 gatttagaatctgcgaatttgcacccatcagcattcagatggtatctgtaacacaa 60
Oy 931 cctgacgacagctcttccagctcggatcagatcattatcagaggtagaagatgtcaa 990
    |||||
Db 61 cctgacgacagctcttccagctcggatcagatcattatcagaggtagaagatgtcaa 120
Oy 991 cagctatcgtcactaactcagatcagatcagatcttggagagagacccctatgtcc 1050
    |||||
Db 121 cagctatcgtcactaactcagatcagatcagatcttggagagagacccctatgtcc 180
Oy 1051 atttgaacaaaagagctcctcacaacacacgcatccgagcaaaattgagtcataag 1110
    |||||
Db 181 atttgaacaaaagagctcctcacaacacacgcatccgagcaaaattgagtcataag 240
Oy 1111 cccagaagactatctcagctgtaactcatttgccttaattgcatctgtcagaagaa 1170
    |||||
Db 241 cccagaagactatctcagctgtaactcatttgccttaattgcatctgtcagaagaa 300
Oy 1171 gtccacatgagggcgaatttgatataatttccagagatgtgcacaaacccacagat 1230
    |||||
Db 301 gtccacatgagggcgaatttgatataatttccagagatgtgcacaaacccacagat 360
Oy 1231 gtcaagctacaaaataacatcatatatagttcaaaaatcggacacaaatacaaaa 1290
    |||||
Db 361 gtcaagctacaaaataacatcatatatagttcaaaaatcggacacaaatacaaaa 420
Oy 1291 ggaacaaagtagcagctcattcttggaaaataatggtatctccgcttccaatgaa 1350
    |||||
Db 421 ggaacaaagtagcagctcattcttggaaaataatggtatctccgcttccaatgaa 480
Oy 1351 tgcctactttgatalagaagagagatcagatcagatgaatttgcacactcgaacagtt 1410
    |||||
Db 481 tgcctactttgatalagaagagagatcagatcagatgaatttgcacactcgaacagtt 540
Oy 1411 aatagtgtaattcctgtgagatctggccagagaagacctggaaactttgacacttcagca 1470
    |||||
Db 541 aatagtgtaattcctgtgagatctggccagagaagacctggaaactttgacacttcagca 600
Oy 1471 ccatttctatacaagagaacatacatcatg----- 1504
    |||||
Db 601 ccatttctatacaagagaacatacatcatg----- 660
Oy 1505 ----- 1504
    |||||
Db 661 tcgatatacaaaattccctggtgatmaaacatcgatgattccttcttgatgaca 720
    |||||
Oy 1505 ---gcactgggattgtacccctcacaatagtgtgtgtgtagaccttaacactgagat 1561
    |||||
Db 721 gaaacactgggattgtgacccctcacaatagtgtgtgtgtagaccttaacactgagat 780
Oy 1562 ggaacagagactagaagcctatcactgagatctgacaacttcccaatccagca 1621
    |||||
Db 781 ggaacagagactagaagcctatcactgagatctgacaacttcccaatccagca 840
Oy 1622 tcagaagttcctgagatgagatgacactcagaagaagtgtagatgacata 1670
    |||||
Db 841 tcagaagttcctgagatgagatgacactcagaagaagtgtagatgacata 889

```

RESULT 2
BE891252

```

LOCUS BE891252 926 bp mRNA EST 20-OCT-2000
DEFINITION 60143077F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3917299 5',
mRNA sequence.
ACCESSION BE891252
VERSION BE891252.1 GI:10350399
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 926)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-femail.nih.gov
Tissue Procurement: ATCC/DC/DTP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9743 row: 3 column: 20
High quality sequence stop: 636.
FEATURES
Location/Qualifiers
source 1..926
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3917299"
/clone_11b="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: skin; Vector: PCMV-SPOrE; Site: 1; Note:
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT 271 a 205 c 212 g 238 t
ORIGIN
Query Match 33.7%: Score 611.2; DB 11; Length 926;
Best Local Similarity 94.0%: Pred. No. 1.3e-137;
Matches 679; Conservative 0; Mismatches 38; Indels 5; Gaps 4:
Oy 87 gaagttccttaagccccaatcctaaagcaagaactgatatgtcagttgtaactat 146
    |||||
Db 1 gaagttccttaagccccaatcctaaagcaagaactgatatgtcagttgtaactat 60
Oy 147 tctgaccagacaagaatgtaaaactaacttgcctgcctcttaagggaaactagaagccct 206
    |||||
Db 61 tctgaccagacaagaatgtaaaactaacttgcctgcctcttaagggaaactagaagccct 120
Oy 207 tccaataattataaaatgagatattgttcgcttccacagagctgaagatccaagtata 266
    |||||
Db 121 tccaataattataaaatgagatattgttcgcttccacagagctgaagatccaagtata 180
Oy 267 taaaagagagactcagggtatccacagcctcctggtccttgaacttgaagttgagggaac 326
    |||||
Db 181 taaaagagagactcagggtatccacagcctcctggtccttgaacttgaagttgagggaac 240
Oy 327 ttggagagccctcctatcactcgcactcgaacgaagattttaaacttaactcagagga 386
    |||||
Db 241 ttggagagccctcctatcactcgcactcgaacgaagattttaaacttaactcagagga 300
Oy 387 ccacaaaatgtagaagcctttagtcttggagcatcactatgacacgtctctgagac 446
    |||||
Db 301 ccacaaaatgtagaagcctttagtcttggagcatcactatgacacgtctctgagac 360
Oy 447 attactaaaattgtgtagatgacagcaatgcaatttgaacctgagctctt 506
    |||||
Db 361 attactaaaattgtgtagatgacagcaatgcaatttgaacctgagctctt 420

```


Accession	Sequence	Position
D6	CTGATTATCTAGATATTTAAATAATTTTAAATTTTAAATTTTAAATAAATAACT	261
Oy	1179 tgaaggcgatttgatataacttcttcgaatggtgcactaaaccccgatgctaacg	1238
D6	2614 ATACGACATTAAGATATTTTAATCTTTAATCTTGATCTGAAAAACACATCTATAAACT	2677
Oy	1239 acaaatatcatatataatgatcctcaaaaatctggaccactaaaalcaaaaagagcga	1296
D6	2674 AGATAAAAAGTTATTCGATTAAGATATAATATATGATGAACGATGAAAAATTTGGAAACGA	2733
Oy	1399 agt--agcgatcattcttgctgaaataatgatgatctccgcttccaatgatgtcta	1356
D6	2734 AGTGATAACGCTAATTTTTTTCATCTTAATAATAGATATTATTTATGAAAGTCAGAACTTTA	2792
Oy	1357 ctttgatagaaggaaggtacaccgcagyaatttgcgaactctcgacaagt	1408
D6	2794 CATTAGATTAATTAATTAAGTCGATTTTTTACCTTTGTTAATAATCAAAATATGT	2845

RESULT 14
 US-08-413-118-72
 : Sequence 72: Application US/08413118
 : Patent No. 5688920
 :
 : GENERAL INFORMATION:
 : APPLICANT: PAOLETTI, ENZO
 : APPLICANT: LIMBACH, KEITH J.
 : TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF
 : NUMBER OF INVENTION: CANINE HERPESVIRUS gB, gC, AND gD AND USES THEREFOR
 : NUMBER OF SEQUENCES: 128
 :
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: CURTIS, MORRIS & SAFORD, P.C.
 : STREET: 530 FIFTH AVENUE, 25TH FLOOR
 : CITY: NEW YORK
 : STATE: NEW YORK
 : COUNTRY: UNITED STATES OF AMERICA
 : ZIP: 10036
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30
 :
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/413,118
 : FILING DATE: 29-MAR-1995
 :
 : CLASSIFICATION:
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/220,151
 : FILING DATE: 30-MAR-1994
 : ATTORNEY/AGENT INFORMATION:
 : NAME: FROMMER, WILLIAM S.
 : REGISTRATION NUMBER: 25,506
 : REFERENCE/DOCKET NUMBER: 454310-2670
 :
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (212) 840-3533
 : TELEFAX: (212) 840-0712
 :
 : INFORMATION FOR SEQ ID NO: 72:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 3659 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 :
 : MOLECULE TYPE: CDNA
 :
 : US-08-413-118-72

Query Match	1.8%	Score 33.2	DB 1	Length 3659
Best Local Similarity	42.7%	Pred. No. 8		
Matches	227	Conservative	0	Mismatches 303; Indels 2; Gaps 1
QY	879	atcgcgaatttcacagcgaatcagcattcagatgttatctgtgcaatcagaacctgaaga	938	
DB	2314	ATCTGATATCTTCTATATCAATTTGGATATCTCCGATTATTCGCCGCTTAAGTAATAATAAA	2373	

Oy	929	cagcttccaaagcttggatcagtcattatatacaggagtagaagaatgttcaaacgctatc	998
Db	2374	TTATTAACTATATATATGTAAATATGATATAGCACTTGCGAAATGGAAATCTGATA	2433
Oy	999	tgtacaatacttaccagatcatcagfatcttggagaagaccactatgtgccatttggaa	1058
Db	2434	TGACAGCAAGATAAATAGATGCTGTGATAAATTAAGTACGAGCATACAAATGATGAATAGACG	2493
Oy	1059	acaaaagctccctcaacaataccgcataccgagcaaatgtgagtcataataagccagaag	1118
Db	2494	AAATATATATATTTGTTTAAATTAATTAAGAAAAATATATTTCTAACTATTCGTTATCAA	2553
Oy	1119	actatttcagtcgttgaacttcatcttgcctaaatgtccatcttgcgcagaagtccaca	1178
Db	2554	CTGAAATATCTAGATATTTAAATTAATTTTATAATTTTATAATTTTAAAAATATCA	2613
Oy	1179	tgaaggcgaatttgataataatttccagaatgtgtcacaataaacccagaatgtcaagct	1238
Db	2614	ATAACGCATTAAGAAATTTTATATCTTTAAATCTTGATCTGAAACCAACATCTATTAACCT	2673
Oy	1239	acaaatatacatataatataigtatccaataatcttgaccctataaaatcgaaaagcgcaa	1298
Db	2674	AGATTAATAAGTTATTCATTAACATATATTAATGCAATGCAACGATGAAAAATTTGGAAACAGA	2733
Oy	1299	agt--agcgacttcatcttggtaaaataatgfatcttccgcgcttccaatgfatgtcta	1356
Db	2734	AGTGAATAGCTTAATTTTTTTCATCTTAAATAGATATATTTTATTGAAAGTACCAAGTTTTA	2793
Oy	1357	cttttgtagaagaaggtacaccctcagtgaaatttgcgaactctgagaacaagt	1408
Db	2794	CGTTAGTAATTAATTAAGCTGCATTTTTCCTTTGTTAATATCAAAATATGT	2845

RESULT 15
 US-08-224-657-48
 ; Sequence 48: Application US/08224657
 ; Patent No. 5756102
 ; GENERAL INFORMATION:
 ; APPLICANT: Paolletti, Enzo
 ; APPLICANT: Tartaglia, James
 ; APPLICANT: Taylor, Jill
 ; TITLE OF INVENTION: POXVIRUS - CANINE DISTEMPER VIRUS (CDV)
 ; TITLE OF INVENTION: RECOMBINANTS AND COMPOSITIONS AND METHODS EMPLOYING THE
 ; TITLE OF INVENTION: RECOMBINANTS
 ; NUMBER OF SEQUENCES: 122
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Curtis, Morris & Safford, P.C.
 ; STREET: 530 Fifth Avenue
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/224,657
 ; FILING DATE: 06-APR-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Frommer, William S.
 ; REGISTRATION NUMBER: 25,506
 ; REFERENCE/DOCKET NUMBER: 454310-2550
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 840-0712
 ; TELEFAX: (212) 840-3333
 ; TELEX: 425066 CURTMS
 ; INFORMATION FOR SEQ ID NO: 48:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3659 base pairs
 ;

```

: APPLICANT: Muller-Cohn, Judy
: APPLICANT: Stamp, Lisa
: TITLE OF INVENTION: No. 6204435e1 Pesticidal Toxins and Nucleotide
: TITLE OF INVENTION: Sequences Which Encode These Toxins
: NUMBER OF SEQUENCES: 134
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
: STREET: 2421 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: US
: ZIP: 32606-6669
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/960,780
: FILING DATE: 30-OCT-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/029,848
: FILING DATE: 30-OCT-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Saliwanchik, David R.
: REGISTRATION NUMBER: 31,794
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 352-375-8100
: TELEFAX: 352-372-5800
: INFORMATION FOR SEQ ID NO: 20:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1045 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: INDIVIDUAL ISOLATE: 31J2
:
: US-08-960-780-20
:
: Query Match 1.88; Score 33.2; DB 4; Length 1045;
: Best Local Similarity 49.48; Pred. No. 5.2;
: Matches 86; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
:
: QY 1285 caaaaggagcgaagtagcagttcatttggtaaaataatggtatctccgcttca 1344
: ||||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
: Db 245 CAAAAGGCCAAAGAAACAAGTGTTCATTAGAAAAAGATAATTAGTCCCATCAA 304
:
: QY 1345 aatgaatgctactcttggatgaagaggtacactcagtgaaattggcaactctgaac 1404
: ||||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
: Db 305 ATTGAATATCAATCTGTGATTAAGCGTTAAACCCAGATAGTCAAAATGTTAAAGAAATTGAAA 364
:
: QY 1405 aagttaatagtgtaattccctgtagatctgcacagaagacccgggaactttg 1458
: ||||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
: Db 365 TTATTTAATAATATAGTCAAAAACAATCTCAGCAAGTGCACAAGACGAATTG 418
:
: RESULT 11
: US-09-073-898-20
: Sequence 20, Application US/09073898
: Patent No. 6242669
: GENERAL INFORMATION:
: APPLICANT: Feltelson, Gerald S.
: APPLICANT: Schnepl, H. Ernest
: APPLICANT: Narva, Kenneth E.
: APPLICANT: Stockhoff, Brian A.
: APPLICANT: Schmeits, James
: APPLICANT: Loewer, David
: APPLICANT: Dullum, Charles Joseph
: APPLICANT: Muller-Cohn, Judy

```

```

: APPLICANT: Stamp, Lisa
: APPLICANT: Morrill, George
: APPLICANT: Finstad-Lee, Stacey
: TITLE OF INVENTION: No. 624269e1 Pesticidal Toxins and Nucleotide
: TITLE OF INVENTION: Sequences Which Encode These Toxins
: NUMBER OF SEQUENCES: 144
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
: STREET: 2421 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: US
: ZIP: 32606-6669
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/073,898
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/029,848
: FILING DATE: 30-OCT-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Sanders, Jay M.
: REGISTRATION NUMBER: 39,355
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 352-375-8100
: TELEFAX: 352-372-5800
: INFORMATION FOR SEQ ID NO: 20:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1045 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: INDIVIDUAL ISOLATE: 31J2
:
: US-09-073-898-20
:
: Query Match 1.88; Score 33.2; DB 4; Length 1045;
: Best Local Similarity 49.48; Pred. No. 5.2;
: Matches 86; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
:
: QY 1285 caaaaggagcgaagtagcagttcatttggtaaaataatggtatctccgcttca 1344
: ||||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
: Db 245 CAAAAGGCCAAAGAAACAAGTGTTCATTAGAAAAAGATAATTAGTCCCATCAA 304
:
: QY 1345 aatgaatgctactcttggatgaagaggtacactcagtgaaattggcaactctgaac 1404
: ||||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
: Db 305 ATTGAATATCAATCTGTGATTAAGCGTTAAACCCAGATAGTCAAAATGTTAAAGAAATTGAAA 364
:
: QY 1405 aagttaatagtgtaattccctgtagatctgcacagaagacccgggaactttg 1458
: ||||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
: Db 365 TTATTTAATAATATAGTCAAAAACAATCTCAGCAAGTGCACAAGACGAATTG 418
:
: RESULT 12
: US-08-220-151-72
: Sequence 72, Application US/08220151
: Patent No. 5529780
: GENERAL INFORMATION:
: APPLICANT: Paolelli, Enzo
: APPLICANT: Limbach, Keith J.
: TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF
: TITLE OF INVENTION: CANINE HERPESVIRUS gb, gc AND gd AND USES THEREFOR

```

1 ADDRESS: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
2
3 STREET: 1100 New York Ave., N. W., Suite 600
4 CITY: Washington
5
6 STATE: DC
7
8 COUNTRY: USA
9
10 ZIP: 20005-3934
11
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE: Floppy disk
14 COMPUTER: IBM PC compatible
15 OPERATING SYSTEM: PC-DOS/MS-DOS
16 SOFTWARE: Patent Release #1.0, Version #1.30
17
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/08/836, 325
20
21 FILING DATE: 2-MAY-1997
22
23 CLASSIFICATION: 514
24
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER: PCT/US95/14251
27
28 FILING DATE: 02-NOV-1995
29
30 PRIOR APPLICATION DATA:
31 APPLICATION NUMBER: 08/482,401
32
33 FILING DATE: 07-JUN-1995
34
35 PRIOR APPLICATION NUMBER: 08/334,029
36
37 APPLICATION DATA:
38 FILING DATE: 02-NOV-1994
39
40 ATTORNEY/AGENT INFORMATION:
41 NAME: Ludwig, Steven R.
42
43 REGISTRATION NUMBER: 36,203
44
45 REFERENCE/DOCKET NUMBER: 0917, 02400002
46
47 TELECOMMUNICATION INFORMATION:
48 TELEPHONE: 202-371-2600
49
50 TELEFAX: 202-371-2540
51
52 INFORMATION FOR SEQ ID NO: 14:
53 SEQUENCE CHARACTERISTICS:
54 LENGTH: 6404 base pairs
55 TYPE: nucleic acid
56 STRANDEDNESS: both
57
58 TOPOLOGY: both
59
60 MOLECULE TYPE: DNA (genomic)
61
62 US-08-836-325-14

Query Match	Similarity	1.9%	Score 33.8	DB 3	length 6404
Best Local	Similarity	53.4%	Pred. No. 7.6		
Matches	71	Conservative	0	Mismatches	62
				Indels	0
				Gaps	0
OY	379	actggagaccacaaatggtagaagccctacgttggtgcacatactatagtcacg	438		
Db	2353	ACTGAGGAAATTCAAAATCTACTTGTATAGGAATTTGGCTTTACGGAAATCTTTGCA	2412		
OY	439	tcctggacattactaaaattgtgtatgcttcagccacatgcattttgacctgactgt	498		
Db	2413	GCTGAATTCGATTATTAATACTGCATTCGATGATCCCATATGATATTTCCAACTAGGCTGG	2472		
OY	499	cagctcttcgggaca	511		
Db	2473	AAATATTTTGACA	2485		

RESULT 8
 US-08-960-780-24
 : Sequence 24, Application US/08960780
 : Patent No. 6204435
 : GENERAL INFORMATION:
 : APPLICANT: Peletelson, Jerald S.
 : APPLICANT: Schmeff, H. Ernest
 : APPLICANT: Narva, Kenneth E.
 : APPLICANT: Stockhoff, Brian A.
 : APPLICANT: Schmeffs, James
 : APPLICANT: Loewer, David
 : APPLICANT: Dullum, Charles Joseph
 : APPLICANT: Muller-Cohn, Judy
 : APPLICANT: Stamp, Lisa
 : TITLE OF INVENTION: No. 6204435el Pesticidal Toxins and Nucleotide

Oy 1285 caaaaggacgaaaagttagcagttcatlctgtgaaaaaatatggtattctcccgcttca 1344
||||||| | ||| , |||||

Query Match	1.9%;	Score 34.8;	DB 4;	Length 1046
Best Local Similarity	50.0%;	Pred. No. 1.8;		

OY 269 TSYGRIAYLE---SNSDVOLK-----DLESANLTANQHSVICOSEPDSPFGVCS 320
DB 417 TRTNGPLVTEELHSLSFEIQLCPGLVLETTSLPV-----VVI-----SNVS 462
OY 321 LRPQMSVARSRLIA 337
DB 463 QLPQMSVARSRLIA 479

RESULT 14
US-08-852-091-4

; Sequence 4, Application US/08852091
; Patent No. 5883228
; GENERAL INFORMATION:
; APPLICANT: James E. Darnell, Jr.
; APPLICANT: Zilong Wen
; APPLICANT: Curt M. Horvath
; APPLICANT: Zhong Zhong
; TITLE OF INVENTION: FUNCTIONALLY ACTIVE REGIONS OF SIGNAL
; TITLE OF INVENTION: TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION (STAT) PROTEINS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,091
; FILING DATE: 06-MAY-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/369,796
; FILING DATE: 06-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 750 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-852-091-4

Query Match 4.7%; Score 84; DB 2; Length 750;
Best Local Similarity 23.9%; Pred. No. 3.8;

Matches 47; Conservative 28; Mismatches 62; Indels 60; Gaps 10;

OY 183 VMDGTRTPPSMRVLIOD-----LVLEGDLSHIRLO-----N 215
DB 301 LMDRT---FSLFOQLIOSSFVERQPCMPHPRPLVLTGVOFTYKRLLVKLOELNIN 357
OY 216 LITDILVYDNHVARSLKAGSLRIYSLATKLSNSENQMLSEF-HL-----HGG 268
DB 358 LKXKVL-FDQVNERVTIVGFRFNILGITHKVMNMEESTNGSLAEFRHLQLEKQKNG 416
OY 269 TSYGRIAYLE---SNSDVOLK-----DLESANLTANQHSVICOSEPDSPFGVCS 320
DB 417 TRTNGPLVTEELHSLSFEIQLCPGLVLETTSLPV-----VVI-----SNVS 462

OY 321 LRPQMSVARSRLIA 337
DB 463 QLPQMSVARSRLIA 479

RESULT 15
US-08-820-754-4

; Sequence 4, Application US/08820754
; Patent No. 5976835
; GENERAL INFORMATION:
; APPLICANT: Darnell Jr., James E.
; APPLICANT: Schindler, Christian W.
; APPLICANT: Fu, Xian-Yuan
; APPLICANT: Wen, Zilong
; APPLICANT: Zhong, Zhong
; TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN
; TITLE OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/820,754
; FILING DATE: 19-MAR-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,185
; FILING DATE: 11-MAR-1994
; APPLICATION NUMBER: US 07/980,498
; FILING DATE: 23-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/854,296
; FILING DATE: 19-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO US93/02569
; FILING DATE: 19-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/126,588
; FILING DATE: 24-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-073 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 750 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-820-754-4

Query Match 4.7%; Score 84; DB 2; Length 750;
Best Local Similarity 23.9%; Pred. No. 3.8;

Matches 47; Conservative 28; Mismatches 62; Indels 60; Gaps 10;

OY 183 VMDGTRTPPSMRVLIOD-----LVLEGDLSHIRLO-----N 215
DB 301 LMDRT---FSLFOQLIOSSFVERQPCMPHPRPLVLTGVOFTYKRLLVKLOELNIN 357

TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 712 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-948-547-6

Query Match 4.7%; Score 84; DB 3; Length 712;
Best local similarity 23.9%; Pred. No. 3.5;
Matches 47; Conservative 28; Mismatches 62; Indels 60; Gaps 10;

OY 183 VMDGTRPPSWAVLIQD-----LVLEGDLSHHRLO-----N 215
DB 301 LMDRT---FSLFOOLIOSSFVERQPCMPTHQRPVLTKGVQFTVKRLVLKLOELNIN 357
OY 216 LTIDILVYDNHVAASLAVGSLRIYSLHTKLSNSSENOTLSLEF-HL-----HGG 268
DB 358 LKVKVL-FKDVNERNTVAGFRKFNILGTHIKVMNNEESTNGSLAEFRHLQLEQKNAG 416
OY 269 TSYGRGIRVLP-----SNSVDOLK-----DLESANLTANQHSVDVLCQSEPDSPFGVYS 320
DB 417 TRNEGRLVTEELHLSLSEFTQLCQPGVLDLETTSLPV-----VYI-----SNVS 462
OY 321 LRPFGMSVARSRLIA 337
DB 463 QLPFGWASILMYNMLVA 479

RESULT 10

PCT-US95-17025-6
Sequence 6, Application PC/TUS9517025
GENERAL INFORMATION:
APPLICANT: James E. Darnell, Jr.
APPLICANT: Zilong Wen
APPLICANT: Curt M. Horvath
APPLICANT: Zhong Zhong
TITLE OF INVENTION: FUNCTIONALLY ACTIVE REGIONS OF SIGNAL
TITLE OF INVENTION: TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION (STAT) PROTEINS
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/17025
FILING DATE: 28-DEC-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/369,796
FILING DATE: 06-JAN-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-116
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
LENGTH: 712 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-17025-6

Query Match 4.7%; Score 84; DB 5; Length 712;
Best local similarity 23.9%; Pred. No. 3.5;
Matches 47; Conservative 28; Mismatches 62; Indels 60; Gaps 10;

OY 183 VMDGTRPPSWAVLIQD-----LVLEGDLSHHRLO-----N 215
DB 301 LMDRT---FSLFOOLIOSSFVERQPCMPTHQRPVLTKGVQFTVKRLVLKLOELNIN 357
OY 216 LTIDILVYDNHVAASLAVGSLRIYSLHTKLSNSSENOTLSLEF-HL-----HGG 268
DB 358 LKVKVL-FKDVNERNTVAGFRKFNILGTHIKVMNNEESTNGSLAEFRHLQLEQKNAG 416
OY 269 TSYGRGIRVLP-----SNSVDOLK-----DLESANLTANQHSVDVLCQSEPDSPFGVYS 320
DB 417 TRNEGRLVTEELHLSLSEFTQLCQPGVLDLETTSLPV-----VYI-----SNVS 462
OY 321 LRPFGMSVARSRLIA 337
DB 463 QLPFGWASILMYNMLVA 479

RESULT 11

US-08-276-099A-12
Sequence 12, Application US/08276099A
Patent No. 5591825
GENERAL INFORMATION:
APPLICANT: McKnight, Steven L
APPLICANT: Hou, Jinzhao
TITLE OF INVENTION: INTERLEUKIN-4 SIGNAL TRANSDUCERS AND
TITLE OF INVENTION: BINDING ASSAYS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276,099A
FILING DATE: 15-JUL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard Aron
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59451-1/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 740 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-276-099A-12

Db 463 OLPGSMASILMYNMLVA 479

RESULT 6

US-08-820-754-6

Sequence 6, Application US/08820754

Patent No. 5976835

GENERAL INFORMATION:

APPLICANT: Darnell Jr., James E.

APPLICANT: Schindler, Christian W.

APPLICANT: Fu, Xian-Yuan

APPLICANT: Men, Zilong

APPLICANT: Zhong, Zhong

TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN

TITLE OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/820,754

FILING DATE: 19-MAR-1997

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/212,185

FILING DATE: 11-MAR-1994

APPLICATION NUMBER: US 07/980,498

FILING DATE: 23-NOV-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/854,296

FILING DATE: 19-MAR-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO US93/02569

FILING DATE: 19-MAR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/126,588

FILING DATE: 24-SEP-1993

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-073 CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201 487-5800

TELEFAX: 201 343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 712 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-820-754-6

Query Match 4.7%; Score 84; DB 2; Length 712;

Best Local Similarity 23.9%; Pred. No. 3.5;

Matches 47; Conservative 28; Mismatches 62; Indels 60; Gaps 10;

QY 183 VMDGTRTFPSMRVLIOD-----LVLEGDLSIHRLQ-----N 215

DB 301 LMDRT---FSLFQOLIGSSFVERQPCMPHTRPQRLVLTGKGYCTVATRLRLVKKQELNYN 357

QY 216 LTIDILVYDNHVAHARSLKAGSEFLRYSLHTKLOSNNSENOTMUSLEF-HL-----HGC 268

Db 358 LKAVVL-FDKDVNERNTVKGFRKFNILGTHTKVMNNEESTNGSLAAEFRLQLKEQKNAG 416

QY 269 TSYRGIRVLP-----SNSVDQLK-----DESANLTANQHSVDVYCOSEPPDDSPNGVS 320

Db 417 TRNEGPLVITEELHSLSFETOLCQPLVYIDLETTSLPV-----VYI-----SNVS 462

QY 321 LRPPGSSVARSRLIAA 337

Db 463 OLPGSMASILMYNMLVA 479

RESULT 7

US-08-956-652-6

Sequence 6, Application US/08956652

Patent No. 6013475

GENERAL INFORMATION:

APPLICANT: Darnell Jr., James E.

APPLICANT: Schindler, Christian W.

APPLICANT: Fu, Xian-Yuan

APPLICANT: Men, Zilong

APPLICANT: Zhong, Zhong

TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN

TITLE OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/956,652

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/212,185

FILING DATE: 11-MAR-1994

APPLICATION NUMBER: US 07/980,498

FILING DATE: 23-NOV-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/854,296

FILING DATE: 19-MAR-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO US93/02569

FILING DATE: 19-MAR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/126,588

FILING DATE: 24-SEP-1993

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-073 CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201 487-5800

TELEFAX: 201 343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 712 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-956-652-6

Query Match 4.7%; Score 84; DB 3; Length 712;

Best Local Similarity 23.9%; Pred. No. 3.5;


```

Db 247 CNGLSMHOVLKRC-----IVGFHLK----- 270
QY 104 LPEFTLGAPIIPRTSSKYFNFTEDHKVLEAL-----VNAS----- 141
Db 271 TLBEGCCGEYIVHSEHGFKLDVESHAIIOVKLLQELLDWKIMNAEGCWETLLAA 330
QY 142 -----THMSPSWTLKLCVOPMOYFDLTQOLLK-----AE 173
Db 331 LEVLIRVHHOOQFNIKOLLNAHVHFLTCVLOEHRREGOLTSMPREVCRSFKILAE 390
QY 174 VDCG-----SFLKLVWDGTRT----- 189
Db 391 VLGSPPDELLVLIFENFLAVHPRTNVCHNPTNFYSLHIDGKIFDEKVOSLAVLRHS 450
QY 190 -----PFPSKVL-----ODLVLEGDSLHHRLOMLTIDILVYDNHVV 228
Db 451 SSGGAPFPSPGLVISPFAFTAPEGTSSSNIVPQMAQMVNSRLP-----AFPTLTP 506
QY 229 VASLKVGSFLRIYSLHFKLOS-----NSENOTMLSLFHLHGTSYGCIGIVLPESNSDV 285
Db 507 LIRAKTLAASIG-FSV-DKIONIADANFEKONLGRPALK--TSKEAFIISCESAKTV 562
QY 286 DOLKKDESANMLIAN-----OHSDVICOSEPPDSFPNGVSL--RP---PGW 326
Db 563 CEMEA-LIGAHASANGVSRGSPFPRAKVDHKDVGTERSDDSPGDESTIPRPDNLKGL 621
QY 327 SSVASRLIAS 338
Db 622 ASFORSOSTVAS 633

```

RESULT 2

```

US-09-021-323-1
: Sequence 1, Application US/09021323
: Patent No. 5929033
: GENERAL INFORMATION:
: APPLICANT: Tang, Y. Tom
: APPLICANT: Corley, Neil C.
: TITLE OF INVENTION: EXTRACELLULAR MUCOUS MATRIX
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Dr.
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/021,323
: FILING DATE: Filed Herewith
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0477 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-855-0555
: TELEFAX: 650-845-4166
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 510 amino acids
: TYPE: amino acid

```

```

: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: SINTNOT13
: CLONE: 1805338
US-09-021-323-1

```

```

Query Match 4.88; Score 85; DB 4; Length 2285;
Best Local Similarity 20.7%; Pred. No. 1.6;
Matches 69; Conservative 48; Mismatches 121; Indels 96; Gaps 16;

```

```

QY 30 KFFRPYLSKTDYCSVYTIYDQNVKLTCLPSGNTEALPIYKNGDIVRHLKIQYV 89
Db 112 KFEKE--LSKREYVQLISYVEKLLNLTVRI--DIMENDISYTELD---FELIRVEVA 164
QY 90 KKETOGITSSGFALEPTEGLGAPIIPRTSSKYFNFT-----TEDEKVVLEALRWAS 142
Db 165 EMERLVIO-----LKESEFGS--SEIVDQLEVLERNMTLLVEKLETIDKNNVLAIR-----R 214
QY 143 HMSPSWTLKLCVOPMOYFDL-----TCOLLKAEVD-----GASFLLKW 184
Db 215 EIVALKTKLKECEASKDQNTPVVHPPTPCSCGCVNVISKPSVYQLMMRGFSYLGAM 274
QY 185 DGRTP-FPS---WRVLIDLVLEGDSLHHRLOMLTIDILVYDNHVAASLKY----- 235
Db 275 GROYSPQHPNKGKLYW---VAPLNTDGRLELYRLVNTLDDLLYLIN---ARELRITYGQ 327
QY 236 GSFRLIYSLHFKLOSNSENOTMLSLFHLHGTSYGCIGIRVLPESNSDVQOLKKLESA 295
Db 328 GSGTAVYNNMVMYNTGN-----IARY 351
QY 296 NLTANQHSVDVICOSEPPDSFPNGVSLRPPGWSV 329
Db 352 NLTTN--TIAVYOTLPMAVNNRFSYANVAWQDI 383

```

RESULT 3

```

US-09-308-375-2
: Sequence 2, Application US/09308375
: Patent No. 6300117
: GENERAL INFORMATION:
: APPLICANT: Genencor International, Inc.
: TITLE OF INVENTION: Proteases from Gram-Positive Organisms
: FILE REFERENCE: GC394-PCT
: CURRENT APPLICATION NUMBER: US/09/308,375
: EARLIER FILING DATE: 1999-05-14
: EARLIER APPLICATION NUMBER: EP9719636.4
: NUMBER OF SEQ ID NOS: 3
: SOFTWARE: FASTSEQ for Windows Version 3.0
: SEQ ID NO 2
: LENGTH: 2285
: TYPE: PRT
: ORGANISM: Bacillus subtilis
US-09-308-375-2

```

```

Query Match 4.88; Score 85; DB 4; Length 2285;
Best Local Similarity 19.88; Pred. No. 18;
Matches 70; Conservative 47; Mismatches 114; Indels 122; Gaps 17;

```

```

QY 1 MSLVPAT-----NYITPLNLQKGTIVNVYGVKFFKPPYLSKTDYCSVYTIYDQIN 54
Db 1464 MQLMPATAKSLGVNNAYDPQNVMGCT-----KYLADOLE----- 1498
QY 55 VKLCLFLSGNTEALPIYK--NGDIVRHL-----KIDV-YKRETOGITSSG 100
Db 1499 -----KFGNVEKALAAYNAGPQNVIAKYGIPPEFKETQYVKKIMANYSKSLSSATSSI 1552
QY 101 FASLTFEGTIGAPIIPRTSSKYFNFTEDHKVLEALRWASHTMSPSWTLKLCVOPMO 160
Db 1553 ASYITNNSAF-----RVSSKY-----GOESGLR--SSPHKGTDFAKAGTAIKSLD 1597

```

